CAMBRIDGE

国外大学优秀教材——微电子类系列（翻译版）

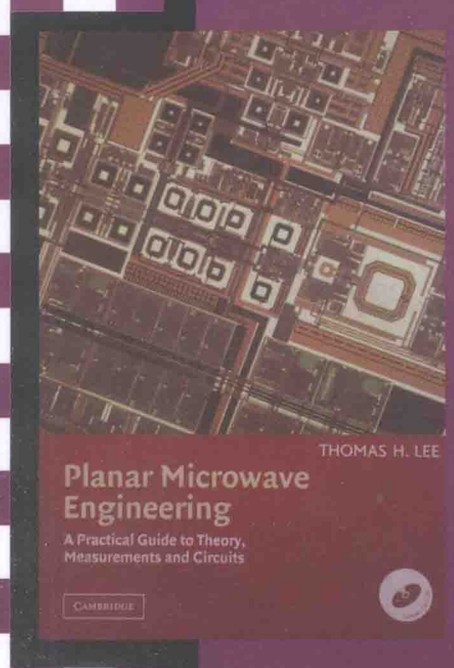
平面微波工程：理论、测量与电路

Planar Microwave Engineering

A Practical Guide to Theory, Measurements and Circuits

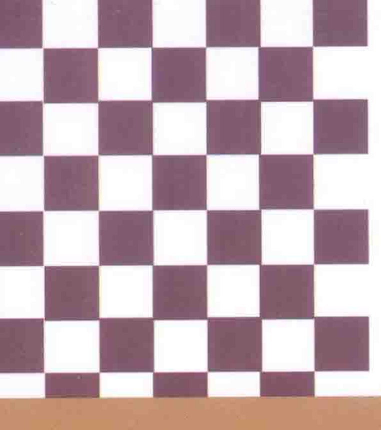
[美] Thomas H. Lee 著

余志平　孙玲玲　王皇　译





清华大学出版社



清华大学出版社数字出版网站
WQBook 书文局泉
www.wqbook.com

CAMBRIDGE
UNIVERSITY PRESS
www.cambridge.org


ISBN 978-7-302-38791-6
9 787302 387916 >
定价：89.00元

国外大学优秀教材 —— 微电子类系列（翻译版）

平面微波工程：理论、测量与电路

Planar Microwave Engineering

A Practical Guide to Theory, Measurements and Circuits

[美] Thomas H. Lee 著

余志平　孙玲玲　王皇　译

清华大学出版社

北京

图书在版编目(CIP)数据

平面微波工程：理论、测量与电路/(美)李(Lee，T. H.)著；余志平，孙玲玲，王皇译. --北京：清华大学出版社，2014

书名原文：Planar microwave engineering: a practical guide to theory, measurements and circuits

国外大学优秀教材. 微电子类系列：翻译版

ISBN 978-7-302-38791-6

Ⅰ. ①平…　Ⅱ. ①李…　②余…　③孙…　④王…　Ⅲ. ①微波电路—高等学校—教材　Ⅳ. ①TN710

中国版本图书馆 CIP 数据核字(2014)第 291256 号

责任编辑：邹开颜　赵从棉
封面设计：傅瑞学
责任校对：王淑云
责任印制：李红英

出版发行：清华大学出版社
　网　　址：http://www.tup.com.cn，http://www.wqbook.com
　地　　址：北京清华大学学研大厦 A 座　　**邮　　编**：100084
　社 总 机：010-62770175　　**邮　　购**：010-62786544
　投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
　质 量 反 馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
印 装 者：清华大学印刷厂
经　　销：全国新华书店
开　　本：185mm×260mm　　**印　　张**：39.5　　**字　　数**：955 千字
　（附光盘 1 张）
版　　次：2014 年 12 月第 1 版　　**印　　次**：2014 年 12 月第1次印刷
印　　数：1～3000
定　　价：89.00 元

产品编号：036442-01

献给安捷莉娜（Angelina）

作者简介

托马斯·李(Thomas H. Lee)在麻省理工学院(MIT)获得科学博士学位,目前是斯坦福(Stanford)大学的教授。他曾经是国际电机电子工程师协会(IEEE)固体电路分会和微波理论与技术分会这两个分会的杰出讲师(distinguished lecturer)。他在国际会议上四次评得"最佳论文"奖,并得到过 Packard 基金会的奖励。李教授发表了超过 100 篇的专业技术文章,并是著名的《CMOS 射频集成电路设计》(The Design of CMOS Radio-Frequency Integrated Circuits)的作者。这本书现在已出了第二版。他持有 35 个美国专利,并与他人创建了几个高科技公司,包括"矩阵半导体"(Matrix Semiconductor)公司。

译者序

美国斯坦福大学教授 Thomas H. Lee 这本《平面微波工程：理论、测量与电路》可以看作是该作者于 1998 年出版(2004 年第二版)的领 CMOS 射频电路教科书之先、在世界范围内十分有影响的《CMOS 射频集成电路设计》(The Design of CMOS Radio-Frequency Integrated Circuits, Cambridge University Press)(以下简称"Lee 射频")一书在微波领域的扩展。

这本英文原版书有 862 页，附有微波教学软件(如 SonnetLite)光盘的宏著，内容涵盖了传输线、天线与滤波器分析、设计及对阻抗、噪声的测量方法。为了对微波电路的知识有一个完整的介绍，本书相当大一部分(大约在全书的 23 章中占 11 章)重复了"Lee 射频"一书中对具有超外差结构的射频收发器的单元(包括低噪声放大器、混频器、振荡器、频率综合器以及功率放大器)电路工作原理的讨论，但阐述的角度更为强调微波的背景及应用。

由于近年来 CMOS 技术的长足进步(现在已处于 22nm 工艺节点)，采用 CMOS 实现的射频电路频率不断提高，从用于手机通信的几百 MHz 到 2GHz，及无线局域网(Wi-Fi)的 2.4/5GHz 频段，逐渐发展到 60GHz，目前更向亚太赫兹(大约在 100～300GHz)的电路前进。本书的中文译本出版，顺应了这方面的科学研究、教学与工程化的需要，相信能在众多的微波书籍中凸显其出色的历史回顾及对电路、系统的深入直观理解，帮助读者以较小的努力进入微波，尤其是 CMOS 平面电路设计的领域。本书可以作为研究生或大学本科高年级学生的射频、微波课程的参考教科书。

本书的翻译历经 4 年，是清华大学与杭州电子科技大学诸多教师与学生(包括研究生与当时的本科生)通力合作的结果。在此特别对清华大学的叶佐昌老师，王洪瑞、冉秋实、张明、杨东旭、袁泽、康佳昊、于欣馨、梁爽、黄声希、郑宏达及赵亮同学，及杭州电子科技大学的刘军与文进才老师，以及苏国东同学表示深切的感谢。没有他们的贡献，此书难以完成。全书的校对、定稿工作由署名作者完成，也因此对可能存在的翻译不准确及笔误负全责。

译者蒙清华大学出版社邹开颜编辑的鼎力支持，从立题到结稿，她以极大的耐性与细致的指导，伴随着我们圆满结束了全书的翻译。在此一并致谢。

余志平(清华大学、杭州电子科技大学)
孙玲玲、王皇(杭州电子科技大学)
2013 年新春伊始
2014 年 12 月付印前

前言

起始，它被称为无线，然后是无线电。无线这个名词在几十年的沉寂后，又变得时髦了。不管我们选择如何去称呼它，射频设计领域的变化是如此之快，教科书的作者们（更不用说工程师了）被逼得不断地跟踪这个领域的发展。对于这个领域中的新手的一个重要挑战是在掌握几代人以来一直没有什么变化的基本知识的同时，吸收指数式增长的新的信息。使得这个挑战更复杂化的是一方面许多有关微波工程的教科书十分偏重电磁场理论，而不讨论那些有实际应用的物理实现例子；另一方面则是一些类似于菜谱的书，几乎完全不涉及理论。更为糟糕的是许多有关这个专题的知识，只是承继了代代口口相传的传统，还往往不正确。其他的有关知识则散落在数不清的应用注解、产品手册、给业余爱好者看的杂志与用户手册中。这些资料中，有许多很难找到，而且互相之间也并不一致。业余爱好者总是不满意学术型教科书过多地强调理论的倾向（“太多的方程，而一直到最后也不告诉你如何去实现”）；而学生及从业工程师们则对给业余爱好者看的杂志的菜谱式的叙述方法不感兴趣（“它们不介绍必要的理论，而通过这些理论则可以将现有的设计改造成一个人所想要的”）。这本书可以满足那些抱怨缺乏平衡理论与实践的带有时效性的参考书的学生、业余爱好者及从业工程师的要求。

这本书是一个在斯坦福大学讲授单学期研究生有关 GHz 收发器设计的高等实验课程——EE414 讲义基础上经过较大扩展的版本。尽管如此，本书的目标不局限于只是那些博士学位课程，而是更大范围的读者面。在 EE414 这门课中，修课的学生们用将近 9 周的时间，用微带线的构建技术，设计、制作与测试一个 1GHz 收发器中的各个模块。这些电路模块包括各种天线与微带线滤波器，以及低噪声放大器、混频器、基于 PLL 的频率综合器与 FM（调频）调制/解调器和功率放大器。这门课程的最后“测验”是一个成功采用这些 FM 收发器的双向通信系统的演示。我深深地特别感谢这门课的第一批学生，他们乐意而热忱地参与这门课的初次尝试，尽管课程课件与讲义的准备在最好的情况下也只是临时抱佛脚完成的。现在这本书的内容与安排很多都是基于这些学生的建议。

毋庸回避，读者会在书中发现相当数量的有关测试技术与微波设计的理论讨论，伴随着不少的传递函数与变换，这是必要的。因为博士生培养计划中本来就有（开个玩笑）在精神上折磨学生的环节。但是对那些只想一步直接获得所需的业余爱好者来说，他们可以忽略所有的方程，而将注意力集中在许多实用的重要设计规则或者是诸多的采用不昂贵的元器件及仪器来实现与表征微波电路的方法。本书作了最大的努力来提供定性与物理层面上的有关遇到的方程意味着什么的解释。特别是对那些喜欢在周末做实验的人，他们可以享受包含在许多章节中的业余制作的二极管及不到 10 美元的微波测量系统。那些从业的工程师则可以找到如何从眼下最先进的仪器中提取最靠得住的数据的方法，特别是注重理解校准方法（以及这些方法的局限性）。这样做就可以避免犯不经意的但却是非常常见的错误。更加年轻的工程师也会感到高兴来发现平时他们对射频（RF）有关话题的疑问的答案（比

如，BNC 是否真的代表“婴儿 N 头的连接器”(baby N connector)?[①] 为什么所有的场合下都用 50Ω? 谁是史密斯(Smith)? 为什么他发明了圆图?)。读者可以自己挑选符合他们口味的话题。这本书就是瑞典式自助餐。这从下面对各章的简略的描述可以清楚地看出来。

第 1 章提供了有关射频与微波电路的一个简史。即使是相对全面地介绍这个宏大的题目，在一章内完成也是不可能的。因此，我们并不试图去这么做。我们的希望是这一章能以轻松的方式为全书的其余部分提供一个背景框架。

第 2 章介绍一些定义与基本概念。试图给出一个不同于在许多其他书籍提供的简单基于频率的有关“微波”的定义。这个定义比较起来不那么具有任意性。我们也尽力避免在书中的任一处出现实际求解麦克斯韦(Maxwell)方程组。代之以采用带有物理直观的方法。再一次强调，这本书是有关设计与测试，而不是纯粹的分析，所以进行的分析都是服务于这些目标的。

史密斯圆图与 S 参数是所有经典微波电路设计的主要成分，所以本书在第 3 章中提供了一个有关它们的简略介绍。史密斯圆图看上去的复杂表象让许多本来对射频与微波设计感兴趣的读者觉得沮丧。因此为了使其不那么吓人，我们通过介绍史密斯圆图的历史来说明它为什么这么有用。

在高频下，功率增益是很难实现的。因此对每一个微波工程师来说，阻抗匹配是一个不可缺少的任务。第 4 章提供了一些阻抗匹配的方法，同时简略解释了 Bode-Fano 极限。这个极限告诉我们什么时候需要放弃阻抗匹配的尝试(或者说，什么时候我们根本不应该去尝试这样做)。

在看到了一些性急的学生不在意地损坏了昂贵的接插件之后，作者感到有必要提供一章专门来讨论如何小心摆弄这些接插件。第 5 章回顾了许多常用的接插件，它们的历史起源与应用范围，以及如何恰当地保养它们。这一章也讨论了电缆线与它们的特性。

第 6 章研究了在微波频率下的集总无源元件的特性。提出了简单的电路模型(适合用于设计，但不一定足够精确到能进行分析)以提醒学生们注意那些方面，它们可能会导致实现了的电路不能按照设计意图进行工作。另外不要轻视在射频与微波设计中的一个重要教训是总是存在着一些不能忽略的寄生元件。设计者必须学会怎样去利用这些寄生元件作为电路的一部分，而不是感到无可奈何。保持对寄生元件的简单而又准确的模型，能使我们设计出达到利用寄生元件为电路服务的聪明的办法。

在第 7 章中，我们介绍制作微波电路(不论是基于分立元件的还是以集成形式)的最常用的方法：微带线。尽管这一章主要讨论这种特定的实现平面电路的方式，我们也花些时间讨论共面波导与共面带状线，以及一般的带状线。在介绍了基本概念与一些简单的(但是相当精确的)设计方程之后，我们考察大量的可以用微带形式实现的无源元件。

一旦已经设计与制作了一些电路，就需要去表征它们。所采用的最基本的测量方法之一是阻抗的测量。第 8 章提供了阻抗测量的几种方法，包括从时域反射计(reflectometry)到矢量网络分析。我们花很大的篇幅描述各种校准技术，因为这里的不留心会导致尽管用了价值 10 万美元的仪器，而只得到值 1 美元的答案。对那些不想在买一个房子(译者注：10 万美元可以在美国买一个房了)与买一个网络分析仪之间作选择的设计者而言，我们提

① 答案是“不是”。

供一个简单的“开槽线”的测量装置，它只需花 10 美元就能得到至少频率高达几个 GHz 的测试功能。幸运的是，这个 40dB(即 10^4，译者注)的成本的节省并不是以 40dB 的使用率下降为代价的。举个例子说，经过校准了的这个装置可以被用来确定频率在 1%～2%以内的精度。

第 9 章专门讨论微波二极管。习惯于低频设计的工程师们经常会惊讶地发现二极管能完成十分广泛的功能。特别令新手感到惊奇的是某些二极管不仅能有放大作用，甚至还能用于产生振荡。

第 10 章在前一章的基础上描述众多的混频器。混频器是现代超外差收发器的核心所在。采用集总或分布的实现都进行了讨论，有源与无源的电路也都在讨论之列。取决于可以利用的技术及设计限制，对于给定的情况，任何上面的一种都可能是适用的。

当然，有源器件是更有趣些(至少对作者来说是如此)，因此第 11 章是有关晶体管的讨论。器件物理工作者近十年来一直在勤奋地进行研究，至今给我们提供了 JFET，MOSFET，HEMT，VMOS，UMOS，LDMOS，HBT，等等。这个器件的清单还在不断变长。我们试图提供对这些晶体管的大概统一的处理方法，并且集中在仅两类(MOSFET 与双极型)的晶体管作为一个更大类的晶体管的代表。

第 12 章考虑如何从我们所具备的晶体管技术中挤出最大的性能来。用几页的篇幅来描述如何对晶体管加直流偏置，因为在微波频率下，许多低频实施技术有其严重的后果(即带来寄生参数，这与前面讨论的是一样的)。然后讨论通过不大的电路复杂度的增加来将带宽展宽 2～3 倍。这里首次(根据作者的知识，此前未见报道)给出详细的对于一个桥接 T 形线圈(T-coil)带宽展宽技术的转移函数与最佳条件的推导。

在第 13 章中，我们将目标由“给我你能提供的所有的带宽”转为“给我最低可能的放大噪声”。首先讨论噪声模型，然后给出了噪声匹配理论。我们发现最大化功率转移的条件几乎不会与最小化噪声因子的条件相重合，因而有必要得到一个折中的策略。又一次，尽管我们集中在讨论一种或二种类型的晶体管，这里给出的一般概念适用于所有的放大器。

一旦已经建造了我们所认为的低噪声放大器，必须证明这一点。第 14 章描述了噪声系数测量所基于的原则，同时描述如何会得到错误的答案(这很容易发生)。取决于最初的目标(是使 LNA 看起来棒呢，还是让竞争对手的产品看起来差)，可以故意犯这些错误或是去避免它们。

第 15 章描述如何得到振荡器所要求的可控不稳定性。在手足无措的微波工程师间的一个陈年笑话是“放大器发生振荡，而振荡器却只起放大作用”。我们希望通过这一章的讨论，使读者能设计真正能起振的振荡器，甚至是在所希望的频率下。

几乎每一个现代收发器在某处都会有一个频率综合器。第 16 章描述锁相环频率综合器，以及有关产生杂散信号的设计缺陷(与如何抑制)的广泛讨论。尽管理论的讨论能变得非常复杂，书中给出的几个设计实例可以帮助缺乏耐心的业余爱好者不需要懂得每一个方程就能完成一个设计工作。

第 17 章分析了相位噪声这一重要的课题。仅仅使得一个振荡器起振是不够的。频谱资源的不足要求每一个发射器遵守“做一个好邻居”的政策，即不在其被指定的频带外发射过多的能量。所有的振荡器在这方面都不是完美的，因此第 17 章确认相位噪声是由哪里来的，以及如何能减小相位噪声。

第 18 章描述相位噪声的测量。就像与噪声系数的测量一样，会有许多不被注意（或是注意到）的方面将相位噪声的测量搞糟糕。我们试图帮助读者梳理清楚这些细节。

第 19 章描述示波器、频谱分析仪与探针。工程师经常只将他们的注意力放在仪器本身，而忘了他们的任务也包括将被测试的电路连接到仪器上。如果在连接时不是足够小心，测量的精度会大大降低。这一章指出了普遍存在的操作错误与避免发生错误的方法。同时，因为高频探针是如此昂贵，本书提供了只需花几个美元就可做一个探针的方法。与探针有关联的，本书也提供了几个可用来测试探针与示波器连接的快速脉冲产生器电路。

第 20 章描述了实现 RF 与微波功率放大器的众多的方法。初看起来，设计满足所要求的增益与输出功率就足够了。遗憾的是，随着电信革命的继续，需求持续的增长，现在设计者必须达到高效率、低成本、应对负载条件变化的鲁棒性以及高线性度。我们试图给出取得所有这些目标的各种方法。

第 21 章的目标是阐述如何将信号功率发射到空气中去，或从空气中取得信号功率。天线看起来总是有点神秘。希望这一章的内容能至少从天线设计中去掉某些神秘感。我们主要讨论微带线贴片天线，但是在开始天线的讨论之前，首先来看一下经典的非平面天线（比如偶极子天线）以便确定一些重要的概念。

最后，第 22 章与第 23 章集中在无源滤波器的设计上。这里的讨论大致分为第 22 章的集总滤波器与第 23 章的微带线滤波器。在这两章的讨论中，我们始终试图给具有实际影响的细节以协调的注意，如元件值容差的效应，或者说是有限的 Q 值。我们希望通过众多的设计实例与模拟结果，会讲清楚设计步骤，让读者很快地将精力集中放在可以被接受与可重复的设计上。

我们再次指出，这些章的组织顺序是半随机的。读者没有必要依书中给出的顺序逐次阅读，可以任意跳过一些章节。记住，这是你自己的书（你有完全控制权）。

这本书的内容吸取了许多令人赞叹的先行者、我的同事与学生提供的信息。他们在过去的日子里与我分享他们的知识与观点。斯坦福的 Malcolm McWhorter 教授（现在是退休的终身教授）主持开发了 EE414 这门课要求的先修课 EE344 中的令人高兴的非常规 BNC 微带线组装步骤。EE344 的中文名称是微波入门实验课程。在这本书的后面会提到，这个组装方法十分适合学生与业余爱好者，因为它使得快速与低代价的制作工作在 GHz 低端频率下的电路原型成为可能。之前在惠普公司工作的 Howard Swain 与 Dieter Scherer 帮助创立了 EE344 这门课，并且一直在帮助教这门课。这两位是惠普公司被广泛使用的微波仪器的天才设计师。EE414 目前的顾问，Donald Cox 教授很慷慨地与我分享他多年来教授 EE344 的经验与体会。这本书的内容经过与他的交流后作了相应的调整。我也要感谢加州理工学院的 David Rutledge 教授，不仅仅是因为他大方地容许在这本书所附的 CD 盘中包括了他主持开发的 Puff 这个程序，而且因为通过与他的个人接触和他发表的著作而给予我丰富的知识。我为 David Leeson 就是我身边的同事这一事实得益匪浅。我的微波系统的知识因为他的随时在旁边而自动地增加了几个分贝。有这么一个如此知识丰富的同事，实在是上帝赐予的礼物。

我也是若干努力工作而又专注的研究生的受益者。他们制作与测试了这本书中描述的大多数课程设计项目。斯坦福的博士生 Sergei Krupenin，Arjang Hassibi，Talal Al-Attar，Moon-Jung Kim，以及 Michael Mulligan 特别应该在此提及。安捷伦的 Rob Chavez，他也

算是 EE414 的一个学生，经常长时间工作来帮助这门课的学生与助教。将来的 EE414 学生以及这本书的读者应该感谢他，因为他的远见卓识与建议成就了目前这些材料的组织。

研究经费是保持学术引擎运行的燃料，在这方面我是十分幸运的。惠普公司与安捷伦技术公司慷慨的仪器捐赠给了斯坦福几代学生直接有最先进的仪器动手的机会，而许多其他学校则没有这个条件。由 William G. Hoover 讲座学者教授基金与 David Packard/Lucile Packard 基金会提供的资助给了我极大的自由空间，使得我可以开发新的课程，探索一些看起来荒唐的研究想法，及撰写教科书。

最后，我深深感谢我所爱的妻子 Angelina，在这本书的写作期间给予我的耐心支持，以及平时的幸福生活。没有她的忍耐，在一个漫长的一年内，完成两本书稿的写作是不可能的。

目　录

第1章 微波技术简史

1.1 引言

许多有关微波技术的历史是从麦克斯韦尔(James Clerk Maxwell)以及他的方程组开始的,而这是有其根本原因的。在1873年,麦克斯韦尔发表了《论电学与磁学》一文,这是他长达十年的试图将两个现象统一起来的努力的巅峰。通过在描述所有先前知道的电磁行为的一组方程中添加一个看起来随意的项("位移电流"),他跨越了已知的真理,而预言了以光速传播的电磁波的存在。反过来说,这个预言不可避免地导致这样一个灼见,即光本身必然是一个电磁现象。学电机工程的学生,或许是被散度、梯度与旋度这些数学运算算子搞晕了头脑,经常不意识到这个内在预言是如何的具有革命性[①]。麦克斯韦尔并不想通过引入位移电流来解决当时没有解决的任何难题。特别是,他没有动机去纠正在电流连续方程中存在的一个明显的电流不连续方程问题(这与许多教科书上给出的标准故事说法正好相反)。他看上去只是被一个美学的感觉而冲动:大自然就是应该让电磁波存在的。不管怎么说,天才这个词,尽管现在有些被滥用了,的的确确适用于麦克斯韦尔的情况。何况天才的英文是genius,而这与精灵(genie)是同出一源的。他所完成的实在是一个奇迹,无疑是19世纪最重要的智慧成就[②]。

麦克斯韦尔,不管是被称为天才还是精灵,在1879去世,才48岁,实在是太年轻了。那一年,赫尔姆·赫兹(Hermann von Helmholtz)设立了一个奖,看谁能率先在实验上证实麦克斯韦尔的预测。1886—1888年间在德国卡尔斯鲁厄工学院(Technische Hochschule in Karlsruhe)进行的一系列出色的实验中,赫尔姆赫兹先前的学生赫兹(Heinrich Hertz)证实了麦克斯韦尔的确是正确的。另一个在这场比赛中的竞争者是洛奇(Oliver Lodge,当时是英国利物浦大学的物理学教授),他在赫兹之后的一个月发表了对麦克斯韦尔理论的证明。在这之前,他刚中断了工作去度假。如果不是这个度假,我们或许今天要用频率单位为兆洛奇等的洛奇波。考虑到Hertz这个字在德语中是指心,而人的心跳大约是每秒一次,看来洛奇没有赢了竞赛倒是一件好事。

赫兹是如何设法用19世纪80年代的设备产生与检测电磁波的呢?实验上的挑战当然要比仅仅产生某种信号要来得大:还需要一个检测器。此外,为了验证波的行为,要求仪器的长度至少为几个波长。这个要求又隐含着另一个条件,即足够的实验室空间来容纳这么大

---

① 数学表述的复杂度甚至会更糟糕。麦克斯韦尔在1873年发表的论文中用一组四元数(quaternion)来描述他的方程组。后来是Oliver Heaviside与Josiah Willard Gibbs(吉布斯,美国数学家,与吉布斯方程是同一人,译者注)抛弃了四元数的用法,而用矢量微积分的语言来重新表述麦克斯韦尔方程组。这样才得到现在这个为大多数现代工程师所熟悉的形式。

② 已故的诺贝尔奖获得者,物理学家Richard Feynman(费曼)经常说未来的历史学家应该赞赏麦克斯韦尔的工作。这是因为与这一发现同一时期发生的另一事件——美国内战,现在看来只具有地区性的重要性。

的仪器(并且理想的情况是这个房间相对于波长而言是无穷大，这样，不想要的从墙壁或其他表面的反射可以忽略)。赫兹当时是一个级别低的教师，他设法搞了一间中等大小的实验室——内部面积大约是12m×8m。[③] 赫兹明白这个实验条件的限制促使他必须寻求产生波长为米量级的信号。他通过实现爱尔兰物理学家 George Francis FitzGerald 的一个猜想，完成了这个任务。FitzGerald 在 1883 年提出了可能可以用当时已经知道的莱登瓶(Leyden jar)(即电容器)振荡火花塞放电来产生电磁波。赫兹意识到这些瓶的半屏蔽结构会妨碍辐射的有效性，他首先修改了 FitzGerald 的主意，将瓶中的圆桶形导体“展平”；然后他用直导线作为电感与这些平板导体连接起来。这样他就可以产生所想得到的频率为几百兆赫兹的谐振频率。在这个过程中，他也顺便发明了偶极子天线。最后，他用一个有缺口的环形天线解决了检测的问题。他的基本发射机—接收机装置见图 1.1。

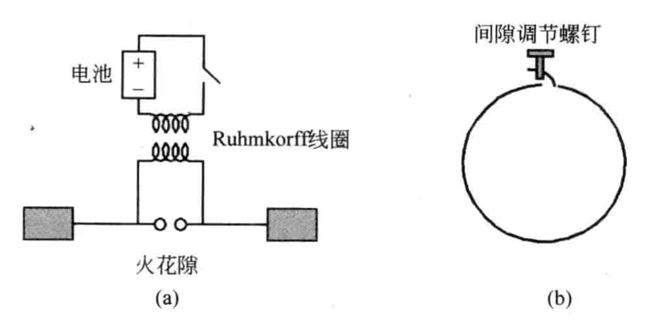


图 1.1 赫兹的火花隙发射机—接收机

当图中的开关合上时，电池给 Ruhmkorff 线圈(一个早期的变压器)的初级绕组充电。当这个开关打开时，磁场的快速消失在变压器的次级绕组上感生出一个高压，从而引发了火花隙的放电。伴随着放电电流的突然变化，激发天线产生辐射。检测过程则是依赖于在环形谐振器中感生出足够的电压来产生可见的火花。一个微米尺度的螺钉使得细微的调整成为可能。在黑暗的环境下观察可以进一步改善测量的灵敏度[④]。赫兹用了这个装置(可以看作是一个光学干涉仪的非常长波长的翻版)演示了基本的波动现象，例如极化与反射[⑤]。波长的测量，以及对电感与电容的解析计算，证实了传播速度足够接近光速。这样，几乎不容怀疑麦克斯韦尔是正确的[⑥]。

我们再也不会知道赫兹是否会在研究这些现象的纯物理内容后，进一步考虑无线技术的实际应用，因为他在 1894 年因为血液中毒(源于感染了的牙齿)去世，才刚 36 岁。看来，我们得在每顿饭后刷牙与用牙线，并且要定时去看牙医。

麦克斯韦尔方程组描述电场与磁场处于一个永不终止的产生、消亡与重生的循环。凑巧的是，麦克斯韦尔的去世激发了赫尔姆·赫兹去设立一个奖项，而这个奖又激发了赫兹。

③ Hugh G. J. Aitken, *Syntony and Spark*, Princeton University Press, Princeton, NJ, 1985.

④ 赫兹也是光电效应的发现人。他注意到在紫外线存在的情况下，火花更容易发生。爱因斯坦因为解释了这个效应而赢得了诺贝尔奖(而不是像通常猜测的那样是因为相对论)。

⑤ 波的反射是比较容易发生的。这个事实激励了不同研究人员在相对一个短暂的时期内提出了导致雷达发明的雏形。

⑥ 这并不是说每个人都是立即被信服了，不是这样子的。革命性的事件需要时间。

赫兹的去世引发了一篇纪念他的文章。这篇文章激发了一个叫做马可尼(Guglielmo Marconi)的年轻人去献身于无线技术的商业应用开发。马可尼是意大利博洛尼亚(Bologna)大学教授 Augusto Righi 的邻居与某个时期的学生。Righi 教授写了纪念赫兹的文章。马可尼出生在一个富裕的家庭,因此他有时间与财力来追求他的梦想⑦。到了 1895 年初,他已经获取了足够多的仪器设备以开始在他们家的郊外住处里外的实验。他努力将通信距离增加。马可尼用了赫兹的发射器,但是不满意火花隙检测器本身的局限。最终他采用(并改造)了一个 Edouard Branly 在 1890 年开发的专门器件。如图 1.2 所示,这个器件,也被洛奇称为相干器(coherer),由一个充满了松散地填入的、可能也被轻微氧化了的金属末的玻璃容器。Branly 偶然发现当附近的电气装置工作时,这个结构的电阻值会发生剧烈的变化。必须强调,相干器工作所基于的详细原理依然不清楚,但是这个知识的缺乏并不妨碍我们来描述它们的电学行为⑧。

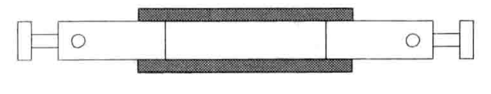


图 1.2 Branly 相干器

一个相干器的电阻在直流工作态下具有大的电阻值(MΩ),而在某个电磁(EM)活动激励下,电阻值会下降到千欧或更少。这一大的电阻值变化被用来触发一个螺线管以产生一个听得见的咔哒声,并打印一条纸带作为被接收信号的永久记录。为了使检波器准备好接收下一个电磁脉冲,必须摇晃它(或"搅拌"它)以恢复成"非相干"的高阻状态。图 1.3 显示了这个检波器是如何被用在一个接收器中的。显然这种相干器是一个数字器件,因而不适合除无线电报以外的应用。

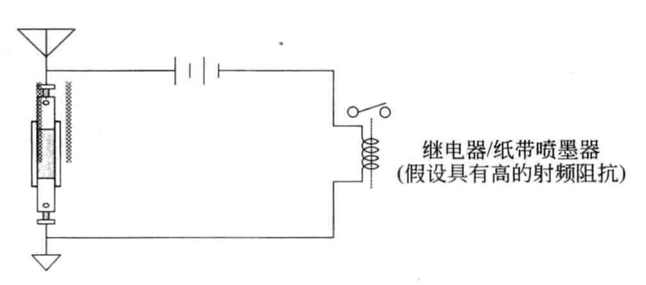


图 1.3 典型的带相干器的接收器

这种相干器从来没有发展成为一个好的检测器,随着时间的流逝,它只是经历了不起眼的改进。马可尼最终停留在如图 1.4 所示的结构。他大大减小了两端塞头之间的距离,在其间的空间充以一种专门的经仔细选择的颗粒大小的镍和银粉的混合物,并且在整个装置

⑦ 马可尼的父亲是一个成功的商人,而他的母亲是 Jameson 爱尔兰威士忌酒庄的继承人。这种家族的联系后来证明是十分宝贵的,在意大利政府对他的发明没有足够兴趣后,它帮助马可尼转向大不列颠政府的关键官员。英国邮政局认可了马可尼的技术,并支持它的后续发展。

⑧ 洛奇将这些器件命名为相干器,因为里面的填充物可以被观察到在某些条件下互相粘在一起。然而,这些器件即使在没有明显可见的物理运动条件下,依然可作为检测器。可以猜测,氧化物的击穿至少是可能的解释之一。但是缺乏对这个解释实验的证据,这是因为对这种器件的兴趣不大。

密封之前将管子抽成部分真空。这一接收器的另外一个改进是在每次接收到一个脉冲之后就自动拍打这个检测器使它回到初始的状态，在这一过程中，接收器中一个附加的螺线管提供了一个可以听得见的信号作为指示。

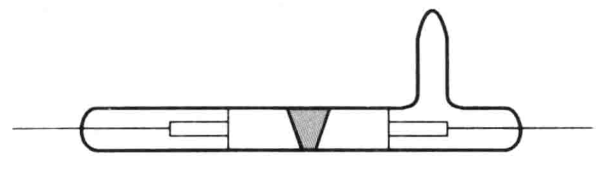


图 1.4　马可尼相干器

尽管所要接收的信号以外其他电磁波事件也能触发相干器，马可尼用这个不怎么理想的器件取得了足够的成功来增加通信距离，在 1896 年达到了大约 3km。当他不断增加功率时，他同时不断地采用更大的天线。这样做所不希望的一个附带效应是载波频率由起先的大约 100MHz 降到了低于 100kHz。这种载波频率的变化是十分幸运的，因为这个低的频率允许电波从电离层反射（当时还不知道有电离层的存在）从而将发射距离延伸至地平线之外。这使得他能在 1901 年 12 月 12 日宣布成功地进行了跨越大西洋的无线通信[⑨]。作为一个结果，其后的无线技术在几乎整整二十年中都忽略了 1MHz 以上的频谱，就是因为相信通信距离在 100kHz 下是最大的。

随着无线电技术的发展，这个相干器的局限性变得越来越不能忍受了。这就引发了探求改进的相干器。没有一个理论来指导好的结构的设计，这种探索往往是无序的，有时甚至会出现荒唐的做法。曾经有人拿刚死的人尸体的脑子来作为检波器，实验者甚至声称他的装置有极好的灵敏度[⑩]。

尽管有这种荒唐的实验，大多数有关相干器的研究是基于一种模糊的感觉，即相干器（或称金属粉末检波器，下称检波器）的工作取决于非理想接触的某种神秘的性质。基于这个直觉的、各种几乎同时进行的对各类点接触晶体检波器的试验都没有成功。这一类器件的第一个专利是在 1901 年由著名的 Jagadish Chandra Bose 申请的。这个专利是关于采用方铅矿（硫化铅）的检波器[⑪]，见图 1.5 与图 1.6。这个检波器利用的是半导体的高电阻温度系数，而不是

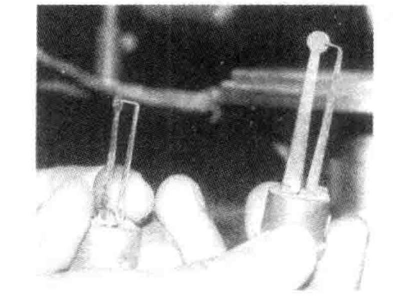


图 1.5　为 Bose 所用的实际的检波器底座（方铅矿没有显示出）
（得到 David Emerson 授权）

⑨　马可尼宣布成功在当时是有争议的，即使在今天依然如此。那个实验本身不是发收双方都是被蒙上眼睛的，因为双方都预先知道发射的信号由字母 s 组成（即莫斯电报码中的三个点）。在他当时采用的长波频段中，无处不在的大气噪声特别强。目前最精确的计算揭示出他当时接收到的三个点一定是噪声，而不是信号。然而，我们不必认为这是一个欺骗。不知情的（即不是故意的）实验者的偏执是一个普遍知道的现象，马可尼的情况也不例外。不管怎么说，马可尼的装置在不到一年内就被改进到足以实现可以被验证的横跨大西洋的通信。

⑩　A. F. Collins, *Electrical World and Engineer*（电气世界与电机工程师），v. 39，1902；他从其他种类动物的脑子开始，一直工作到人的脑子。

⑪　美国专利#755 840，于 1904 年 3 月 19 日批准授予。这个专利持有人的姓名给出的是 Jagadis Chunder Bose（而非 Jagadish Chandra Bose，译者注）。我们采用的译法是与印度加尔各答的那个用他的名字命名的学术研究所相同的。

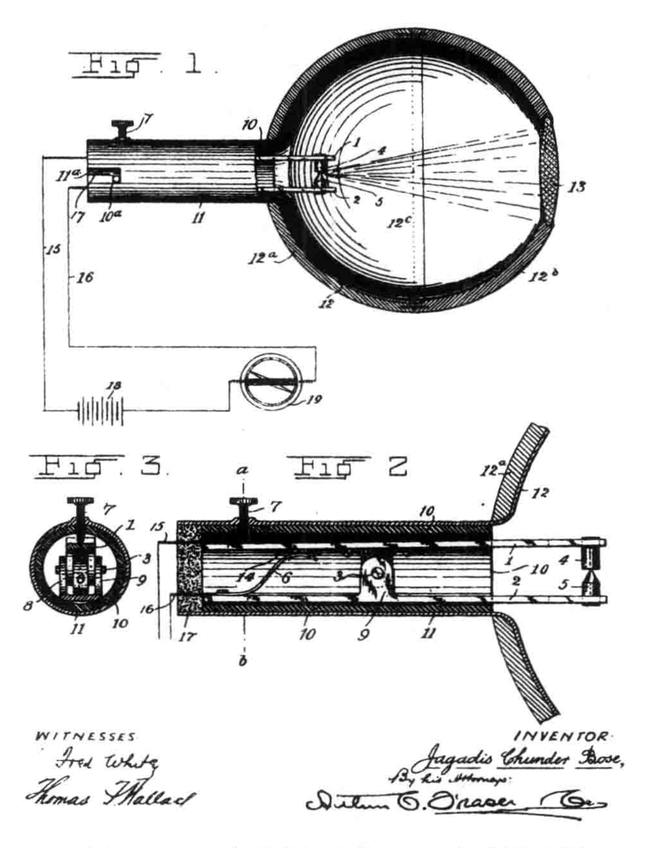


图 1.6 Bose 的热度计(bolometer)专利(第 1 页)

它的整流特性[⑫]。可以从专利的手画图中看出,电磁辐射是聚焦在点接触上。而与这样产生的加热过程相伴随的电阻值的变化被记录成为流过外部电路的电流。这一类检波器被称为热测定器。经改进了的热测定器依然被用作测量功率的一种手段,特别是对那些频率如此之高的信号,以至于没有别的办法可以检测。Bose 用这种检波器将实验的频率推高至大约 60GHz。他在 1891 年的文章中首次发表了这个结果[⑬]。他对微米波现象的研究领先了

⑫ 许多关于 Bose 工作的描述将他的方铅矿热度计(bolometer)与后来发展起来的点接触(猫须状类)整流检测器混淆了,从而错误地将半导体二极管的发明归功于他。后者(即半导体二极管)当然有整流作用,因此不需要外加偏置。早在 1874 年,Ferdinand Braun 就报道了在方铅矿和黄铁矿(以及其他一些矿物质)上存在着不对称的导电特性(Ueber Stromleitung durch Schwefelmefelmetalle(论通过金属硫酸盐的电流),*Poggendorff's Annalen der Physik und Chemie*. v. 153, pp. 556-63)。Braun 的其他一些在无线技术上的重要贡献是将火花隙与变压器的初级线圈串联起来,而这个变压器的初级是连到天线的。他后来因为对无线电技术的贡献而与马可尼共享了 1909 年的诺贝尔物理学奖。

⑬ J. C. Bose, "On the Determination of the Wavelength of Electric Radiation by a Diffraction Grating,"(论用衍射栅决定电辐射的波长),*Proc. Roy. Soc.*, v. 60, 1897, pp. 167-178.

他的时代几十年[14]。Bose 在英国剑桥大学的老师，瑞利(Rayleigh)爵士，认识到中空的导体可以传递电磁能量，这同样是超前了他的同时代人几十年[15]。通过波导来传递信号这样的做法在 19 世纪 90 年代之后的 40 年中都没有人提起过，但瑞利在 1897 年就解决了有关的大部分问题(包括截止频率的概念)。

这个专利似乎是授予半导体检测器的第一个，尽管当时还没有清楚地认识到这一点，因为半导体那时还没有被看作是专门一类材料(的确，"半导体"这个词当时还没有造出来)。沿着这个方向的工作依然在继续进行，Henry Harrison Chase Dunwoody 将军在 1906 年 3 月 23 日申请了第一个采用金刚砂(碳化硅)的整流检波器的专利。这个专利申请在当年的 12 月 4 日得到批准，专利号#837 616。一个在 Dunwoody 之后申请的专利，即 1906 年 8 月 30 日由 Greenleaf Whittier Pickard (一位 MIT 的研究生，他的祖父辈长辈是诗人 John Greenleaf Whittier)就一个硅(!)检波器提出的专利申请，却在 Dunwoody 之前得到批准：1906 年 11 月 20 日，专利号#836 531(见图 1.7)。

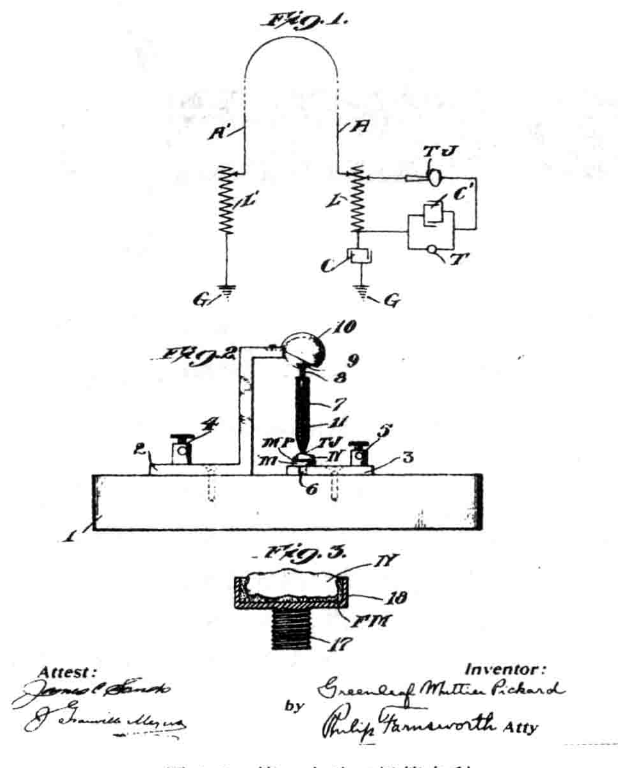


图 1.7 第一个硅二极管专利

⑭ 对 Bose 的毫米波工作的一个极佳的回顾，见 David T. Emerson，"The Work of Jagadis Chandra Bose: 100 Years of MM-Wave Research,"(Jagadis Chandra Bose 的工作：毫米波研究的 100 年)，*IEEE Trans. Microwave Theory and Tech.*, v. 45, no. 12, 1997, pp. 2267-2273.

⑮ 大多数的科学家与工程师熟悉瑞利的关于声学的广泛著作，包括管道(声波波导)的分析。但很少有人知道他也奠定了电磁波导的基础。这是一个时代，当时没有人能想象可以利用这个现象，而且除了 Bose 也没有人甚至能产生频率足够高的波在尺寸那么小的波导中传播。

如图 1.8 所示，一端连接的是一根细导线（奇怪地被称“猫须”），它形成了到晶体表面的一个点接触；另一端连接的则是一个大面积的接触，由低熔点的合金形成（通常是铅、锡、铋和镉的混合物，被称为“Wood”金属，其熔点在 80℃以下）。这个大面积接触包住了晶体[16]。我们可以把这样构成的器件称为点接触式肖特基二极管，尽管测量设置并不总是那么容易地与这一描述相一致。不论怎样，我们可以看到现代的二极管符号是如何从这一实际装置的描述中演变而来的，二极管符号中的箭头代表触须的点接触。

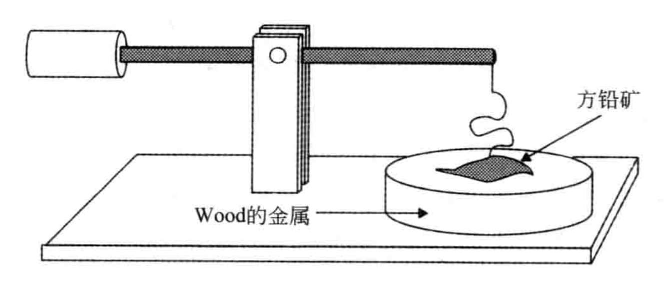


图 1.8　典型的晶体检波器

图 1.9 给出了用这些器件制作的简单的晶体收音机[17]。一个 LC 电路调谐到所要接收的信号，然后晶体对此进行整流，只留下解调了的声波来驱动耳机。某些检波器（如方铅矿）无需直流偏置源，所以有可能制作一个“无需能量”的收音机[18]。就像我们将会看到的，一个在还是小男孩时就被晶体无线电的魔力所迷住的人，长大后会试图用点接触二极管来实现雷达的开发。晶体无线电依然是一群被这种接收器的魅力所吸引的无线电爱好者的强烈兴趣的焦点。

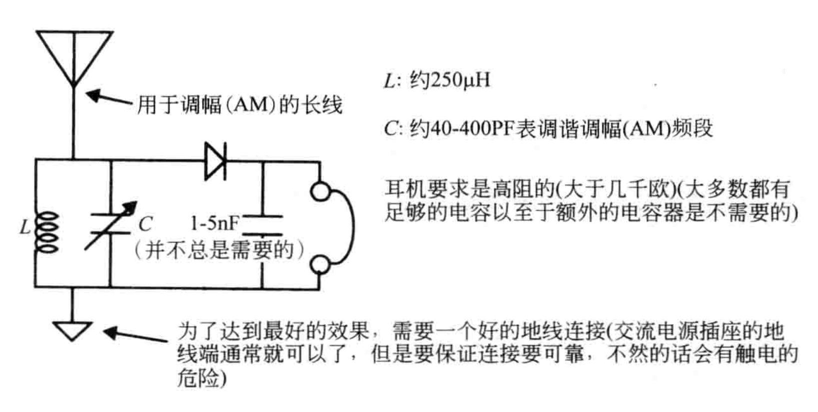


图 1.9　简单的晶体无线电收音机

---

[16] 尽管这么说，这种晶体被接触包住是不必要的。与晶体的体部的紧密压接通常就足够了。而这样做又可以避免使用带有毒性的金属。

[17] 今天，晶体通常是指用在诸如在振荡器中作为决定频率的元件的石英谐振器；这个称谓与用在晶体收音机中的晶体没有关系。一个方铅矿晶体可以用买得到的商用二极管（例如锗 1N34A）替代。但是理想主义者不会批准这样缺乏魅力的做法。一个普通的美国一分钱硬币（不早于 1983 年），在厨房的烘箱中 250℃下烤 15 分钟以形成氧化铜（CuO），具有许多与方铅矿类似的特性（如完全不可预测的行为）。其他国家的用铜做的硬币也可能会工作（作者已经验证了韩国的 10 韩元硬币工作得特别出色）。读者被鼓励用世界各地的硬币做试验，然后告诉作者。

[18] 或许我们应当感谢人类的听觉系统：人类听觉的阈值相当于耳膜的一个氢原子直径那么大的位移！

Pickard在开发晶体检波器时比任何其他人工作得更努力，他最终试验了30 000种导线和晶体的组合。除了硅以外，还有黄铁矿石(傻子的金子)和生锈的剪刀。方铅矿检波器曾非常普及因为它们价格低廉且不需要任何偏置。遗憾的是，触须线接触的恰当调节很难被保持，因为只要在方铅矿上的压力稍微大一点就会破坏其整流性能。此外，还得先在晶体表面搜寻一个敏感点。另一方面，尽管金刚砂检波器需要有几伏的偏置，但它们在机械上比较稳定(较高的接触压力也毫无问题)，因此广泛地应用在船上[19]。

大约在这些原始的半导体初次使用的同时，无线电工程师们开始与由火花信号的宽频谱引起的干扰问题作斗争。这一信号的宽带特性倒是很好地适合相干器技术，因为反正相干器阻抗的剧烈变化已使得调谐电路的实现十分困难。然而火花隙技术不适合多路通信这一问题在1901年独显出来。当时三个独立的组(分别为马可尼、Lee de Forest与Pickard所领导)试图去同时提供美洲杯赛艇比赛的无线实时报道。当这三个组同时发出火花信号时，没有人能收到能够理解的信号。结局是竞赛结果不得不用老办法，即信号机，来报告。气恼之极的de Forest把他的发射机扔到船外，而在岸上急着等新闻的中继台则不得不编造许多他们报告的内容。

作为对这一事件的回答，一些工程师寻求产生具有射频频率的连续正弦波方法。其中之一是极富天赋的丹麦工程师Valdemar Poulsen[20](因为发明过一个早期的磁记录装置而出名)。他采用了与直流弧光放电关联的负电阻特性来保持一个*LC*电路不断地振荡[21]。一个刚刚毕业的斯坦福研究生，Cyril Elwell，确保了Poulsen的电弧发射机的生产许可而在加州帕拉阿图市(Palo Alto，与斯坦福大学邻近，译者注)创建了Federal Telegraph(联邦电报)公司。这家公司很快将这个技术扩展到能输出十分吃惊的功率水平：第一次世界大战之后没有多久，一个功率为1MW的电弧发射机就投入使用了。

Reginald Fessenden要求通用电气公司(GE)的Ernst F. W. Alexanderson设法采用不同的技术途径，用巨大的交流发电机来产生大功率的射频正弦波(这个发电机其实与我们开车时给汽车电池充电的那个一样，只是非常大，速度非常高)。这样一个已经走到了头的技术在制造出一个频率为100kHz、功率达200kW的交流发电机后，也算达到了顶峰。这个发电机在第一次世界大战结束时才完成，但在它开始要工作时，该技术已开始变得落后了[22]。

---

[19] 金刚砂探测器通常被封装在一个盒子里，并且它们的调节常常是通过一个十分小心地将其压在一个硬表面上的步骤实现的。

[20] 有些资料总是把他的名字不正确地写成是“Vladimir”，这完全不是一个丹麦人的名字！

[21] 用于工业照明的弧光技术那时已是十分成熟了。需要有一个足够大的串联电阻来补偿电弧的负阻(因而能维持一个稳定的电流)是十分明白的。William Duddell利用了负阻的特性来产生音频(以及可以听得到的)振荡。Duddell的“能唱歌的弧光”或许能用来作娱乐用，但没有什么其他特别的用处。试图将振荡频率提高到音频范围之上的努力并不成功，直到Poulsen转而用氢气，并且采用了一个强磁场在每个周期里将离子扫除出去(这个主意是Elihu Thompson在1893年提出来的)。Elwell随后将这个装置的尺寸放大以取得更高的功率。这个方法足以将功率增至30kW。但简单的通过这种按比例放大来增加输出功率的方法，只导致了越来越大的发射机。而输出功率依然被限制在30kW上。在Leonard Fuller的博士论文中(这是Stanford电机工程的第一篇博士论文)，他做了理论上的改进。这个进步使得电弧功率能突破上面的壁垒，而实现了1MW的电弧发射机。在1931年，作为加州大学伯克利(Berkeley)分校电机工程系的系主任 — 而且那时候电弧技术已经变成历史了 — Fuller安排了Federal Telegraph(联邦电报)公司将多余的绕线圈的机器与一个80t的磁铁捐赠出去以建造Ernest O. Lawrence(劳伦斯)的第一个大型回旋加速器。劳伦斯用这个装置后来在1939年赢得了诺贝尔物理奖。

[22] 这样一个先进的依赖于旋转子的机器已用尽了冶金技术所能提供的所有潜力，要进一步提高工作频率，比如说高于200kHz，是完全不可能了。

连续波比火花信号更有优势这一点立即就变得很明显,并由此刺激了开发更好的接收设备的努力。幸运的是,相干器逐渐为许多改进了的器件,包括前面描述的半导体器件,所替代。到了 1910 年时,相干器已注定是要被淘汰的了(尽管迟至 50 年代至少有一种无线电控制的玩具还在用相干器)。

到了 1906 年后期,已有足够多的整流检波器投入使用,以至于美国东海岸的船员们吃惊地发现他们能收听到 Fessenden(Reginald Fessenden,无线电话的加拿大发明者,译者注)本人在圣诞节前夜的第一次调幅广播(尽管三天前,这已通过无线电报作了预告)。向兴高采烈的听众广播的节目包括亨德尔(Handel)的广板(Largo)(来自歌剧《赛尔斯》(Xerxes))、Fessenden 用小提琴演奏的圣善夜(O Holy Night)(Fessenden 自己还边拉小提琴边唱了最后一段歌词),以及他发自内心的对听众的圣诞节祝福。[23] 他用了一个水冷的碳精粉麦克风来调制一个 5kW(大约的)、50kHz(也是大约的)载波。这个载波是由位于美国麻省 Brant Rock 这个地方的一个原型的 Alexanderson 交流发电机产生的。那些使用相干器的不幸运听众错过了这一历史事件。Fessenden 在一周之后,即在新年前夜又重复了他的广播,使得更多的人有机会通过参与得到乐趣。

1.2　真空管的诞生

1907 年见证了 Lee de Forest 的第一个能起放大作用的电子器件的发明:真空三极管。遗憾的是,de Forest 并不明白他的发明是如何实际工作的。他是通过来回几次的机遇(有时甚至是缺乏职业道德的方式)才碰巧有这个发明的。

真空管要追溯它的祖先到爱迪生(Thomas Edison)的不那么高科技的白炽电灯泡。爱迪生的灯泡有一个问题,即由于烟垢(碳灯丝的蒸发所致)在灯泡内表面的积累而逐渐变黑。为了解决这个问题,他插入了一个金属电极,希望能把烟垢吸附到这个电极而不是玻璃上。他不愧是一位真正的实验家,对这一电极分别加了正电压和负电压(相对于灯丝的一个接头),在 1883 年注意到当这个电极是正电压时会有奇怪的电流流过,但当这个电极是负电压时则没有任何电流。同时流过的电流大小取决于灯丝的温度。他当时还没有任何理论可以解释这些观察到的现象(注意,"电子"这个词直到 1891 年才被 George Johnstone Stoney"造"出来,并且直到 1897 年汤姆逊(J. J. Thomason)的实验才没有疑义地确认了这个粒子),但爱迪生继续他的研究并在 1884 年获得了第一个电子(而不是电气)器件的专利。这一器件利用板极电流依赖于灯丝温度这个关系来间接测量上的线电压[24]。这一装置从来没有被生产过,因为相比于标准的电压计其性能要差些;爱迪生就是想得到另一个专利,就此而已。这是他最终获得了 1093 个专利的一种方式。

大约就在同一时期,一个叫弗莱明(John Ambrose Fleming)的英国爱迪生公司的顾问碰巧去参加了一个在加拿大的会议。他在美国逗留期间去拜访了他在新泽西州的兄弟,以

---

[23] "An Unsung Hero: Reginald Fessenden, the Canadian Inventor of Radio Telephony,"(一个未被赞颂的英雄:Reginald Fessenden,加拿大的无线电话发明家)(http://www.ewh.ieee.org/reg/7/millennium/radio/radio_unsung.html).

[24] 美国专利#307 031,1883 年 11 月 15 日申请,1884 年 10 月 21 日授权。

及爱迪生的实验室。他为这个“爱迪生效应”大受启发(远比爱迪生要感兴趣，而爱迪生本人很难理解为什么弗莱明会对这个没有任何实际应用希望的现象如此兴奋不已)，并最终在1890—1896年间发表了有关这个效应的几篇论文。尽管弗莱明的实验最初引起了一些轰动，但伦琴(Röntgen)在1896年宣布发现了X射线——以及在同一年的后些时候发现了天然放射现象——之后，爱迪生效应很快便不被人注意了。

几年后，弗莱明成为了英国马可尼公司的顾问并开始寻求改进检波器性能的努力。他想起了爱迪生效应，经过一些灯泡的测试，发现了这些灯泡完全能用作射频整流器。于是他在1905年申请了弗莱明真空阀(Fleming Valve，真空管因此在英国仍被称为阀)的专利(图1.10)[25]。几乎完全耳聋的弗莱明采用了一个镜子电流计(mirror galvanometer)来使接收到的信号以可见的形式表示出来，并把这一特点作为他专利的一部分。

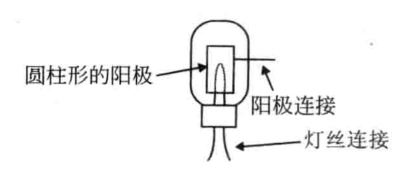


图1.10　弗莱明阀(真空二极管)

尽管并不是特别灵敏，但弗莱明阀至少能连续作出响应并且不需要任何机械的调整。马可尼公司安装的许多设备都采用它们(但多半是由于契约义务的关系)。其实弗莱明阀从来就没有被普遍接受过——这与某些历史的说法正好相反——因为它消耗太大的功率，灯丝的寿命很短，而且这一东西又很贵。当与制作精良的晶体检波器相比时，它的灵敏度又不高。

de Forest此时正忙着在美国悄悄地创建几家无线电公司来与马可尼竞争。他在1902年早期的杂志上充满希望地写道，“相信我们，很快就会有鱼儿上钩的”。的确，他的公司在美国很快就是仅次于马可尼无线通信公司的最大无线通信公司了。de Forest绝不是一个轻易让机会溜走的人，他剽窃了弗莱明二极管并甚至设法在1906年得到了专利授权(#836 070，5月19日申请，11月13日授权)。他只是将弗莱明的镜子电流计换成了一个耳机，然后加上一个极大的正向偏置(这样就将一个原本不灵敏的检波器变得更不灵敏)。证明de Forest直接偷窃了弗莱明的工作的结论性证据最终出现了，一位叫Gerald Tyne的历史学家得到了H. W. McCandless的商务记录，后者是de Forest所有第一批真空管(de Forest称它们为“电子三极管，audion”[26])的生产商。这些记录清楚地显示了de Forest在申请他专利的几个月前曾要求McCandless复制过一些弗莱明真空阀。因此根本不可能有善意的解释来说明de Forest独立发明了真空二极管。

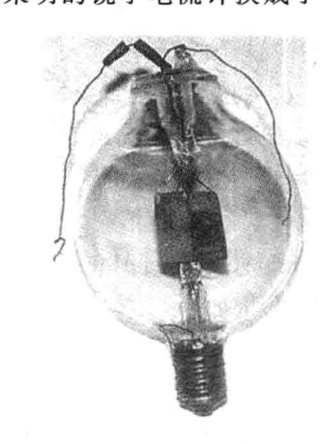


图1.11　De Forest的真空三极管(为Armstrong用在他1912年的再生接收机的双阳极真空三极管(来自Houck收藏，经Michael Katzdorn许可))

然而他的下一个成就倒是他自己的，而且十分重要。他在灯丝和翼电极(后来称为板极)之间增加了一个用曲曲弯弯的导线做成的电极，de Forest把它称为栅电极，这样就诞生了真空三极管(见图1.11)。这一有三个部件的真空三极管有放大的能力，但de Forest直到若干年之后

[25] 美国专利#803 684，1905年4月19日申请，1905年11月7日授权。

[26] Gerald F. J. Tyne, *Saga of the Vacuum Tube*(真空管传奇)，Howard W. Sams & Co(出版公司)，1977.

才认识到这一特点。事实上，他的专利申请描述的三极管只是作为一个检波器而不是一个放大器[27]。所以增加一个栅极的目的仍然是谜一样不清楚。他增加这个栅极肯定不是像某些史料所称的那样是经过仔细考虑后的结果，事实是他到处加电极。他甚至想在板极外面加“控制电极”！因此我们必须把他增加栅极这个行为仅仅看成是他在寻找称得上是他自己的检波器时所作的一种随机但又是持续的修修补补的结果。如果他偶然发明了真空三极管的这个说法是对的话，那么要由别人来向他解释这个结构是如何工作的却是必然的[28]。

根据已有的证据，无论是 de Forest 还是其他任何人都在好几年内没有对真空三极管想的太多(直到 1912 年，这种器件的年销售量都不超过 300 件)[29]。在某个时候，de Forest 在他的公司破产清盘之后，不得不放弃他所有能从他的发明中得到的利益。但只有一个发明是例外的：律师们允许他保留三极管的专利，认为真空三极管没有任何价值。失业并破产了的 de Forest 去了帕拉阿图市的联邦电报公司上班，老板是 Fuller。

面临着几乎没有什么别的选择，de Forest 与联邦电报公司的工程师 Herbert van Etten 和 Charles Logwood 一起工作，进一步发展了真空三极管，并且在 1912 年下半年发现了真空三极管放大作用的潜力。几乎在同时，其他一些人(包括火箭先驱者 Robert Goddard)[30] 也发现了。他在那一年费了劲把这一器件卖给了 AT&T 作为电话的继电器的放大器，同时保留了将这个器件用于无线通信的权力。但最初一个时期他依然很困难，因为三极管性能很不稳定[31]。器件特性的重复性相当差并且管子具有有限的动态范围，它在小信号下能工作得很好，但在过载时性能很差(管子中的残留气体会电离，导致蓝色辉光以及在输出信号中的劈裂噪声)。更为严重的是三极管的灯丝(它是用钽制作的)寿命只有约 100～200h。因此还需要经历一个时期，真空管才会席卷整个世界。

1.3　阿姆斯特朗(Armstrong)与自激放大器/检波器/振荡器

幸运的是，真空三极管的前途并不只在 de Forest 一个人手中。通用电气公司(GE)实验室的 Irving Langmuir 尽力设法获得了更高的真空度，这样就消除了由于存在(容易电离的)残留气体引起的古怪的特性。de Forest 专门警告过要避免有高的真空度，这部分是因为他真的相信这会降低灵敏度，同时也是因为他不得不对他自己与别人维持这样一个神话，

㉗ 美国专利＃879 532，1907 年 1 月 29 日申请，1908 年 2 月 18 日授权。然而足够奇怪的是，他的有二个部件的声频管专利确实包含了放大作用。

㉘ Aitken 在《连续的波》这本书(普林斯顿大学出版社，新泽西州普林斯顿，1985)中争辩说 de Froster 曾被不公平地谴责为不理解他自己的发明。然而一大堆证据都是与 Aitken 宽容的观点相矛盾的。

㉙ 美国专利＃803 684，1905 年 4 月 19 日申请，1905 年 11 月 7 日授权。

㉚ Goddard 的美国专利 1 159 209，在 1912 年 8 月 1 日申请，1915 年 11 月 2 日被批准，描述了一个三极管振荡器，因此实际上甚至比 Armstrong 的有文档记录的工作还要早。

㉛ 尽管他当时是联邦电报公司的正式雇员，他私下与 AT&T 独立地谈判成条款，这样就违犯了他的雇用合同。但联邦电报公司选择没有采取法律行为。

即他的发明与弗莱明的二极管没有任何关系[32]。

Langmuir 的成就为一位聪明的工程师铺平了设计出有用的、利用真空三极管潜力的电路的道路。这位工程师就是阿姆斯特朗(Edwin Howard Armstrong)。他于 1912 年在他稚嫩的 21 岁时发明了自激放大器/检波器[33]。这一电路(它的现代版本如图 1.12 所示)采用了正反馈(通过一个所谓的"板极回馈线圈"将输出能量中的一部分以恰当的相位耦合回输入端),来同时提高系统的增益和 Q 值。于是具有高的增益(这可以取得高灵敏度)和窄的带宽(这可以有好的选择性)的电路就可以相当简单地只用一个管子就能实现。此外,这一管子的非线性也可以被用来解调输入信号。进一步而言,将输出和输入过度耦合就可以使这个电路变成一个极好而又紧凑的 RF 振荡器。

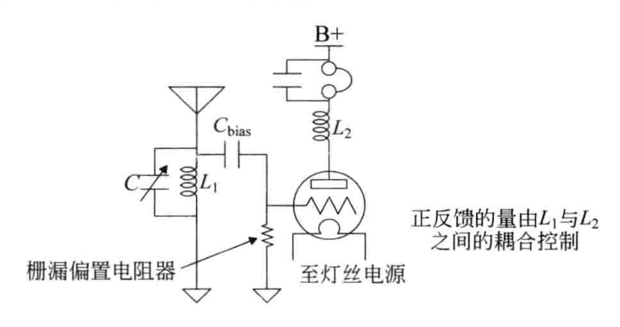


图 1.12 阿姆斯特朗的再生接收机(见美国专利#1 113 149)

阿姆斯特朗在 1914 年发表的"真空三极管的工作特点"[34]的论文中,首次给出了真空三极管是如何工作的正确解释,并提供了大量的实验证据来支持他的论点。随后的另一篇论文《真空三极管接收器的最新进展》[35],解释了再生放大器/检波器的工作,并且阐述了过强的耦合是如何将放大器变成一个振荡器的。这篇论文堪称是一个论述清晰的典范,就是对于现代的读者,其可读性也相当好。这篇文章激怒了 de Forest 的程度可以从该文发表后立即发生的记录在案的他们间的书面交战看出。读者可以读到 de Forest 尴尬地来找出阿姆斯特朗工作中错误所作的软弱无力的试图。在他的气急败坏的反击中,de Forest 犯了许多可笑的错误,表现出对一些基本概念的正确理解都有困难(例如他得出相当于认为正弦波的平均值为非零这样的结论)。这样,最终暴露出他并不明白真空三极管实际上是如何工作

---

[32] de Forest 观察到当他在用他的无线仪器时,他的实验室里的充气灯泡的亮度似乎在发生变化。他就想是否可以利用这个现象来用作检波器。进一步的研究揭示了这些灯泡实际上是对由他的火花发射机产生的声波噪声一种反应。就是基于这样一个不靠谱的线索,de Forest 编造了一个他是怎样被这个尽管是令人失望的"火焰探测器"的实验激发出一个想法,即气体对电磁波是有反应的,并最终导致他独立于弗莱明发明了真空三极管。

作为一个工程师,de Forest 是不够格的,他在遣词造句上倒颇有本事。他通过反复地提及将气体放在一个"剧烈的分子活动的条件"下,竭力来解释火焰探测器的工作原理(见美国专利#979 275)。在他的自传("无线电之父")中,他这样描述了这个类似于相干器的装置的工作过程(但他没有指明他是从 Fessenden 处剽窃了这个装置的结构):"这些微小的摆渡船,每一个都装载了一小点电荷,航行到另一侧的电极将其轻飘飘的以太载荷卸下来"。或许他是希望用这种文学的语言来遮盖这些句子中根本没有任何科学内涵。

[33] 他经过公证的笔记本记载的日子实际上是 1913 年 1 月 31 日,仅仅比 de Forest 的真空三极管能放大的发现晚几个月。

[34] *Electrical World*(电气世界),1914 年 12 月 12 日

[35] *IRE Proceeding*, v. 3, 1915, pp. 215-247

的，而这是他自己的发明(其实只是一个发现而已)。

这两个人之间的严重对立从来也没有减退过。

阿姆斯特朗继续他的电路开发，这些电路直到今天仍然在电信系统中占据主导地位。作为第一次世界大战期间美国陆军通信部队(U.S Army Signal Corps)的一名成员，阿姆斯特朗参与研究远距离探测敌机的问题，他设想利用飞机上的点火系统(这又是火花发射机)自然发出的信号来定位对方。遗憾的是，在大约1MHz频率以下并没有发现多少有用的辐射，而当时的真空管又极难对比这个频率更高的信号放大。事实上，直到1917年H.J. Round经过了非同寻常的努力后才取得了在2MHz频率下的有用的增益，因此阿姆斯特朗并没有完成这个使命。

他后来通过构建一个由Fessenden拥有专利的系统来解决了这个问题。Fessenden寻求通过解调连续波(continuous wave, CW)信号来解决这个问题。在Fessenden的外差式解调器中，一个高速交流电发生器起到本机振荡器的作用，将射频信号转换成音频频率，使得用户可以挑选一个能躲避干扰的音频频率。通过使得从不同发射机发出的信号变成十分容易分辨的音调，Fessenden的外差系统能在干扰存在的情况下，实现前所未有的清晰程度。

阿姆斯特朗决定以不同的方式采用Fessenden的外差原理。他不是用它来直接解调连续波信号，而是用外差方法把一个高频RF的输入信号先转变成一个频率较低但依然是超声频率的信号。在这个频率下可以较为容易地得到高的增益和选择性。这个被称为中频(intermediate frequency, IF)的较低频率的信号，经过了滤波和放大之后再被解调。这样一个接收器可以很容易地达到足够的灵敏度，以至于限制因素实际上就只是大气的噪声(这在调幅广播频段是非常大的)。更进一步的优点是使得采用单一调谐控制成为可能，这是因为IF放大器在固定频率下工作。

阿姆斯特朗在1917年获得了超外差接收器的专利(见图1.13)。尽管在阿姆斯特朗可以采用超外差结构来探测敌机之前，战争就已经结束了，他在几个有才干的工程师(包括他一生的朋友与助手)的帮助下继续开发，最终把真空管的数目从原先总共10个减少到5个(这同样是一件好事：原来对总灯丝电流的要求为10A)。RCA公司的David Sarnoff最终协商购买了这一超外差技术。结果是到了1930年时，RCA已开始主宰无线电市场。

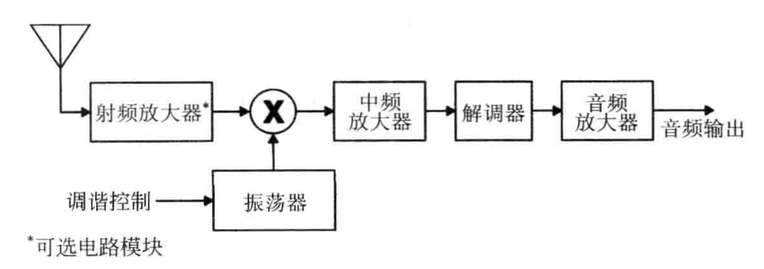


图1.13 超外差接收器的框图

第一次世界大战的需求，加上日益增长的对电话的要求，驱动了真空管及与之关联的电子学的迅速发展。这些进步反过来又成就了无线通信的应用。这个应用可是远远超出了马可尼原来的预见。对称的点对点通信系统。这个通信系统模仿有线电报的工作方式。一旦无线电话的技术出现了，那些后者的初始预见大部分是有眼光的人，如Fessenden与

Herrold“博士”，所作的先驱努力，凸显了无线技术作为单点到多点的娱乐媒介[36]的潜在的商业价值。对这样一个革命性的主意，因为在历史上没有先例，只得被迫从农业上借用一个词来描述它：广播(即撒种子)。无线电广播很快就占据了重要地位，以至于无线的潜力看起来是没有限制的。在20世纪20年代，数百个新成立的无线设备公司出售的接收器充斥了市场。到了20世纪20年代末期，超外差结构已经变得重要了。RCA的股票在当时是市场上的佼佼者，从1924年的大约每股11美元窜到考虑到分股因素的总值每股114美元的高位。这是因为投资者对这个(无线)行业注入了大量的资金。不幸的是，1929年的股票市场大崩溃终于导致RCA的股票随着无线泡沫的吹破而到1932年掉到每股仅3美元。

伴随着无线的快速增长，对稀有的频谱资源的竞争也增加了。这是因为通常使用的频率集中在被认为是最有用的亚1MHz频带。一个三方利益冲突，包括了业余无线电爱好者(“hams”)、政府利益与商业服务，是通过限制业余爱好者到1.5MHz以上的频率部分得到解决的。这部分频率在当时是被认为没有什么用的。没有什么其他可以选择的执着的业余爱好者尽他们的最大努力来应对这个局面。出乎每个人的意料，这些业余爱好者们发现了这个相应于波长为200m或更短的“短波”频谱的巨大的价值[37]。这个成就解放了工程师们的想象力去考虑更高的频率，从而激发了20世纪30年代的思索微波的浪潮。

1.4 奇才间的战争

尽管用于商业目的的无线电广播驱动了在第一次世界大战后大多数无线技术的发展，正是一个不断增长的，对高于几MHz的频谱可能会有用的这种认知，直接导致了性能更好的真空管与先进电路技术的出现。进行电视广播的建议具体化了，用于军事的通信不断地得到快速发展。在同一时期，AT&T开始研究用无线技术来补充他们的电话网络。对更多的频谱的需求变得日益迫切，高频设计的掌握程度很快发展到1MHz以上，先是到10MHz，然后在20世纪30年代中期达到了100MHz。

随着频率的增加，工程师们面临着一系列的困难，其中一个困难是在电缆中大的高频衰减。当意识到同轴电缆中的导体损耗几乎完全是因为中心导体的小的直径，人们自然就想到了这个引起麻烦的中心导体是否需要[38]。这个思路引导了两个组的人员去探索通过中空的管道来传递无线电波的可能性。分别由贝尔实验室的George C. Southworth与MIT的Wilmer L. Barrow这两个组彼此独立地工作，且同时在1936年中宣布了他们的进展[39]。微

---

[36] Charles “Doc” Herrold在那些无线电的先驱中是比较独特的，因为他坚持无线电用于娱乐的发展。在1909年，他开始固定地从位于加州圣荷西(San Jose)市及附近的一群发射机中广播音乐与新闻的节目。这个固定节目一直延续到20世纪20年代，然后这个广播站被出售，并被迁移到旧金山(在那里，它变成了KCBS)。参见《广播事业的被遗忘的父亲：有关Charles Harrold的故事》，1994年KTEH电视台的广播节目。这个节目的脚本可以在http://www.kteh.org/productions/docs/doctranscript.txt找到。

[37] 见：Clinton B. DeSoto，*Two Hundred Meters and Down*(200m及更短)，美国无线电继电器协会，1936年。业余爱好者因为他们这些努力得到的回报却是：在他们证明了这个频谱的有用性后不久，就不允许他们再用这个频段了。

[38] 比如说，Heaviside在19世纪90年代就曾经想过这个问题，但是想不出该如何在没有这第二个导体的情况下来处理(信号)。

[39] 见：G. C. Southworth, High Frequency Waveguides—General Considerations and Experimental Results(高频波导——通用考虑与实验结果)，*Bell System Tech. J.*(波尔系统技术期刊)，v. 15, 1936, pp. 284-309。Southworth与Barrow直到他们要在同一个会议上分表各自的工作一个月前还彼此不认得。他们最初也不知道瑞利(Rayleigh)爵士40年前就已经奠定了相关的理论基础。

波在波导中的低损耗传输很快被证明对一个 Southworth 与 Barrow 都没有预见到的应用是起了关键作用，这个应用就是雷达。Southworth 对高频信号检波器的需要也导致他在他的同事 Russel Ohl 的建议下回到了硅点接触(猫须)检波器。这个半导体的重新启用也对今后产生了一个十分深远的影响。

勉强接受在欧洲战争是不可避免这一事实引发了重新来考虑几十年前提出的建议：雷达[40]。英国人特别有前瞻性，他们是首批采用一个被称为 Chain Home(连锁雷达网)系统在空防中部署雷达的。这个 Chain Home 系统在 1937 年开始运行[41]。一开始的工作频率是 22MHz，随着系统的搜索范围与能力的扩展，频率增加至 55MHz，正好赶上在不列颠战争中起到了关键作用。到了 1941 年，一个工作频率为 200MHz，被称为 Chain Home Low(低空连锁雷达网)的系统已经在运行了。

对于雷达应用而言，更高的工作频率在理论上被证明是非常有利的，但是缺乏合适的检波器与大功率的信号源阻碍了被称为微波的实际应用发展。在当时，微波这个词意味着大约是 1GHz 或更高的频率。普通的真空管存在着基本的按比例缩减的限制，因而使得工作在微波波段上有困难。当频率增加时，电子的有限的速度迫使应用越来越小的电极间的间距，从而减小了管子的功率处理能力。因为功率正比于电压的平方，真空管的输出功率随频率增加呈平方律减小。

在 1937 年，Russell Varian 发明了一种利用渡越时间效应的真空管来避开这些缩减限制[42]。见图 1.14。这个器件是与 Varian 的兄弟 Sigurd Varian 和物理学家 William Hansen 在斯坦福(Stanford)大学一起发明的。速调管首先将电子(这些电子是由加热了的阴极提供的)加速到一个高的速度(比如说，光速的 10%)。这个高速电子束然后通过一个空腔谐振器的多孔的平行栅。加在这些栅上的放置的信号(即输入信号)对进入空腔的电子或者加速，或者减速，取决于加在栅上电压的瞬时极性。一旦离开栅区，这些电子在一个低电场区域内漂移，在其中，速度较快的电子赶上了较慢的电子，形成了周期性的聚堆(bunching)(希腊语中的聚堆是 klystron，因此而得名)。这个从恒定的电子密度到空间脉冲状密度的转换使得电子束的电流中出现了在信号频率下的成分。然后，第二个谐振器选择了这个频率成分(或者希望的话还可以是其谐波)，一个耦合线圈则提供了与外部世界的界面。速调管受渡越延迟效应的影响较小，部分是因为电子是首先被加速(这样对于给定的振荡周期，就容许采用较大的栅极间距)，然后被(输入信号)控制(而在一个普通的真空管中，电子电流的控制发生在一个电子移动缓慢的区域)；部分是因为渡越延时本身是为形成在漂移空间聚堆所必需的。作为一个结果，在微波频率下的特别高的输出功率是可以达到的。

速调管放大器通过简单地提供到输入端的反射就能转换成一个振荡器。这样的反射可以通过有意设计来产生，或者是从第二个谐振器的不希望发生的失配得到。由 Varian 与贝尔实验室的 John R. Pierce 在 1938 年或 1939 年独立发明(实际上是发现)的反射速调管通过将第二个谐振器用一个被称为“排斥极”的电极取代来利用这个对反射的敏感性。反射速

---

[40] 一个经常被引用的例子是 Christian Hülsmeyer 在 1904 年 3 月申请的关于“Telemobilskop”的专利。见 1906 年 1 月 16 日被批准的美国专利#810×510。实际上在这个专利中没有什么新的主意。

[41] 英国对雷达的名称是无线电定向搜索(radio directional finding，RDF)，但这个名字没有被普遍采用。

[42] 美国专利#2 242 275，1937 年 10 月 11 日申请，1941 年 5 月 20 日批准。

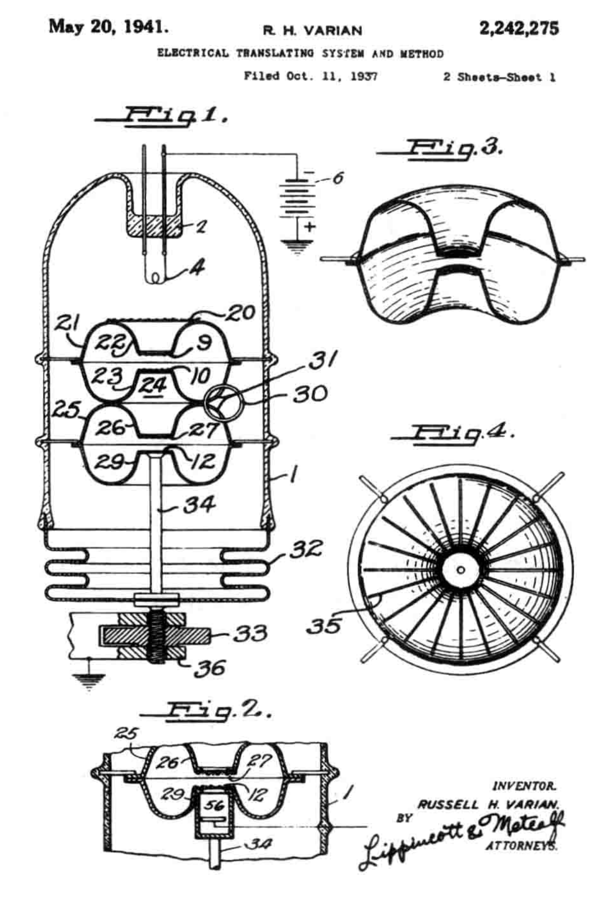


图 1.14　Varian 的速调管(klystron)专利的第一页

调管当时被广泛用作雷达接收器的本机振荡器，因为它们体积小并且相对来说容易被调谐(至少在一个有用的区间内是如此)。

另一个器件，空腔磁控管(cavity magnetron)，逐渐发展成能提供雷达发射机以令人惊骇的输出功率(比如说，脉冲输出功率可达 100kW)。最早的磁控管结构在 1921 年为 GE(通用电气)的 Albert W. Hull 所描述[43]。Hull 的磁控管就是一个阳极为圆筒形的二极管。由一个位于中心的阴极发射出的电子在沿着电子管轴向所加的磁场作用下经由一条弯曲的路径送达阳极。Hull 发明这个交叉场器件(之所以这样称呼是因为电场与磁场是沿不同的方向)的动机与高频的产生是毫无关系的。事实上，通过用磁场(而不是通常的栅极)来控制电流，他只是简单地试图来设计一个不会侵犯现有专利的真空管。

意识到磁控管的潜力远不止只是为了躲避现有的专利这个事实是一个缓慢的过程，但是到了 20 世纪 30 年代中期，寻求可以工作在更高频率下的真空管的试图已经导致了几个独立的小组去仔细研究磁控管。一个例子是贝尔实验室的工程师 Arthur L. Samuel 在 1934 年申请的专利[44]。这个发明与重新兴起的目标专门为高频工作的磁控管的发展不谋而

[43] 见：*Phys. Rev.*(物理评论)，v. 18，1921，p. 31，以及“The Magnetron”(磁控管)，*AIEE J.*，v. 40，1921，p. 715.

[44] 美国专利＃2 063 341，1934 年 12 月 8 日申请，1936 年 12 月 8 日批准。

合。很快地，天才的德国工程师 Hans E. Hollmann 发明了一系列磁控管，其中某些结构与后来 Henry A. H. Boot 及 John T. Randall 在 1940 年建造的空腔磁控管十分类似[45]。

Boot 与 Randall 在英国的伯明翰(Birmingham)大学的工作多多少少是在主流的雷达研究之外进行的。他们的主要任务是开发改进了的雷达探测器。自然，他们需要某些信号来供探测。然而，缺乏合适的信号源促使他们去想可行的办法。对于新近被开发出来的速调管的热情被这个管子的机械复杂度所减弱(的确，第一个速调管是由 Sigurd Varian 所构造的，而他是一个有天赋的机械工程师)。他们决定把精力集中在磁控管上(见图 1.15)，这相对于速调管在结构上相对简单些。在 1940 年 2 月 21 日，Boot 与 Randall 用他们的原型的磁控管验证了第一次微波发射。仅在几天内，他们产生了 3GHz 频率以上的输出功率就达到了令人吃惊的 500W。这个成就几乎是当时所能达到的最先进的功率值的 100 倍[46]。

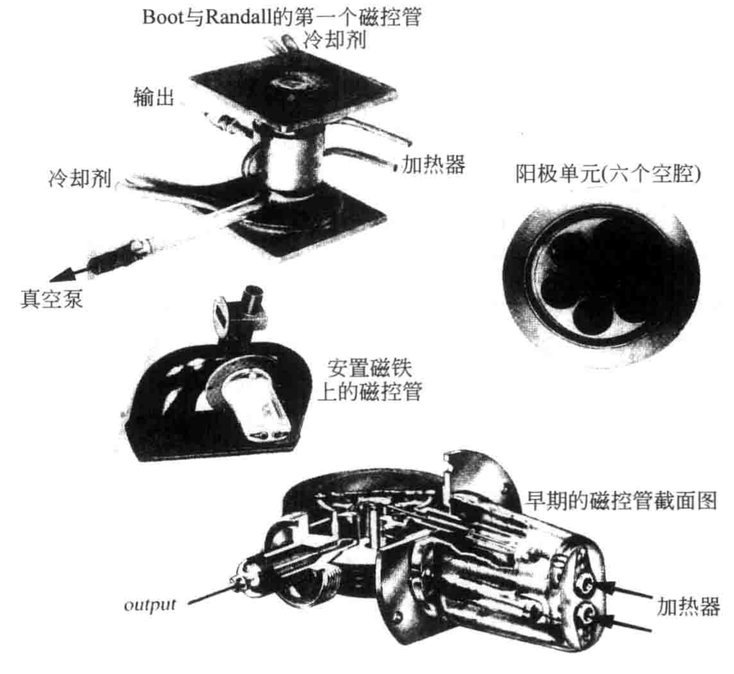


图 1.15　磁控管[47]

磁控管依赖于与速调管同样的堆聚现象。但是在这里静止的磁场引起电子沿着弯曲的路径由中心阴极运动到阳极板。当电子通过那些谐振器时，取决于加在谐振器间隙上的瞬

[45] 美国专利 #2 123 728，1936 年 11 月 27 日申请，1938 年 7 月 12 日批准。这个专利是基于一个较早的在 1936 年申请的德国专利。这个德国专利在当年(1936 年)为 Hollmann 写的书《*Physik und Technik der Ultrakurzen Wellen, Erster Band*》(超短波物理与技术，vol. 1)。Hallmann 将功劳归于 Heinrich Greinacher，而不是 Hull。这个经典的参考引用对战争期间技术的发展在英国与美国产生了远比在德国大的影响。

[46] 就像其他重要的发展，对是谁发明的及什么时候发明的存在着争议。从存档的记录来看，有关磁控管的专利要比 Boot 和 Randall 的工作要早。但是这个记录并不排除与之独立的发明。俄国人可以引用 Alekseev 与 Maliarov 的工作(第一次在 1940 年发表在一个俄国期刊上，然后重新发表在 *Proc. IRE*, v. 32，1944 年)；德国人可以指出 Hollmann 关于这个器件的大量的发表；等等。这种观点显然是与战时的雷达故事不相干，因为正是联盟国单独将这个发明推进到十分重要的程度。

[47] George B. Collins，MIT Rad. Lab, Ser.，vol. 6(微波磁控管)，McGraw-Hill，New York，1948.

时电压，电子或者被加速，或者被减速。就像在速调管中那样，电子的聚堆发生，而谐振器将基波信号取出来。在其中一个谐振器上的耦合线圈提供了到外部负载的输出[48]。

Boot 与 Randall 的空腔磁控管的优良性能对于雷达技术进步的影响程度可以从这样一件事看出：英国的 Tizard（人名）使团于 1940 年 8 月在高度机密，间谍战的气氛中将一个使用磁控管的原型机带到美国 。磁控管使所有亲眼见到它操作的人都十分惊讶，因而 Tizard 回到联合王国（即英国）就带走了当时在“官方”立场上还处于中立的美国人保证进行雷达合作开发的承诺。[49]

磁控管使所有亲眼见到它操作的人都十分惊讶，因而 Tizard 回到联合王国（即英国）时带来了从当时在“官方”立场上还处于中立的美国人的进行雷达合作开发的保证。

Tizard 使团的成功很快就导致了麻省理工学院（MIT）辐射（radiation）实验室的成立。到了 11 月中，Rad Lab（在当时被这么称谓，且延续至今）成了一个正式的实体。最终，它与曼哈顿计划在经费与研究人员上相互竞争。在“二战”的过程中，更多的经费是花在发展雷达上，而不是发展原子弹。

磁控管促成了空中与地面雷达两个方面的惊人进步，许多研究组在一起工作（至少在大部分情况下是如此）。比如说，贝尔实验室与 Rad 实验室既是竞争对手，又是合作伙伴。著名的 SCR-584 瞄准雷达就结合了 Rad 实验室的雷达与贝尔实验室的模拟计算机，使得不仅可以跟踪一个目标，而且能自动瞄准与驱使武器开火。

到了战争结束时，磁控管已可以产生 10GHz 频率的 100kW 脉冲了，而 24GHz 的雷达也在发展中[50]。这种在那么多前沿领域中同时令人屏气的先进技术的快速发展可以通过浏览战后不久为 Rad 实验室的工作人员编纂的 27 卷书中看出[51]。雷达技术很快被采用到从烹煮到无线电天文学的无所不包的应用。Pickard 的硅猫须二极管的进化至复杂的厘米波混频器二极管为晶体管的发明与半导体时代的起始搭建了舞台。可以这么说，现代电子学是起源于第二次世界大战的。

第二次世界大战也见证了无线电技术超越雷达的迅速发展。便携式无线通信的有用性很早就被认可了。1940 年摩托罗拉（Motorola）给美国陆军通信兵（Army Signal Corps）提供了手持调幅收发机的对讲机[52]。到了 1941 年，商用的双向移动调频通信系统就出现了，而其军用的版本则出现在 1943 年（即 15kg 重的 SCR-300 背包式收发机，这是第一个被叫做 Walkie-Talkie 的）[53]。

---

[48] 这个解释显然不够全面，因为留下了这个操作是如何开始的问题。对这个问题的答案是噪声足以使整个过程进行。一旦振荡开始了，这里提供的解释就显得有道理了。

[49] 在第二次世界大战期间，英国的磁控管有 6 个谐振空腔，而美国版本则有 8 个。人们可能会倾向于将这个差别归结于“不是在这里发明的”症结，但是在这里却不是这个解释。英国人只造了一个有 8 个谐振腔的原型机，而这正是被 Tizard 使团随机挑到的，这个原型机就变成了美国磁控管的先祖。

[50] 这最终是因为水蒸气在这个频率附近致使的高大气吸收率而失败了。然而，这个失败却触发微波频谱议的诞生。因此，这个故事最后是有一个圆满的结局。

[51] 如果包括索引在内，那么就有 28 卷。不管怎么说，整套 Rad 实验室的书现在是以 2 张 CD 的形式在 Artech House 出版。

[52] 顺便说一下，缩写 SCR 代表 Signal Corps Radio，手持对讲机（Handi-Talkie）用了一套 5 个真空管工作在一个用晶体来选频的 3.5～6MHz。这个对讲机很快就变成了一个偶像般的标志，出现在数不清的有关第二次世界大战的报纸与电影中。同样地，SCR-584 与它的可旋转天线在许多战后科幻“B”片中用作布景，天线朝着天空扫描来发现所要找的目标。

[53] 曾在第一次世界大战期间在陆军通信兵服过役的、十分爱国的阿姆斯特朗将他的调幅专利无偿让美国政府在第二次世界大战期间使用。

移动通信在战争期间被多次证明是十分有价值的，因此，这种认识自然激发了战后的移动无线通信的发展。密苏里州的圣路易斯成为第一个享受移动无线电话服务的城市。这个服务被恰当地称为 MTS(mobile telephone service，移动电话服务)[54]。其收发器工作在 150MHz 频带，有 6 个信道，信道间有 60kHz 间隔，采用 FDD(frequency-division duplexing，频分双工)(即，一个频率用于上行，一个频率用于下行)以及频率调制[55]。因为不同的频率允许多用户同时通话，这个系统也代表了早期的用在移动无线网络中的频分复用(frequency-division multiple access，FDMA)[56]。

在 20 世纪 50 年代，雷达技术继续得到发展，而且微波炉(专利为 Raytheon 的工程师 Percy Spencer 所拥有)也开始出现了。由磁控管提供微波功率，工作频率为 900MHz，这些早期的微波炉比现在家庭厨房中见到的要大不少[57]。

平面微波电路也在那 10 年随着印刷线路板变得日益普遍而出现。苏联成功地在 1957 年 10 月发射了 *Sputnik* 卫星激起了空间竞赛，并且开始出现微波技术的航空航天的应用。性能更好的微波仪器的广泛使用也促进了无线电天文学的重要的成就，包括开始于大约 1955 年的对宇宙背景辐射的首批测量。然而，这些实验的重要性在很长一个时期没有被认识到，以至于在这个领域给贝尔实验室的 Arno Penzias 与 Robert Wilson 以后赢得诺贝尔物理学奖留下了空白[58]。

随着晶体管在 20 世纪 50 年代与 20 世纪 60 年代早期的进步，它们逐渐在更高的频率下取代真空管，并且这种取代的发生愈加频繁。完美地适用于平面形式的实现以及航空航天的应用，固态放大器逐渐占据了小功率的微波技术。同时，对于在更高频率下性能更好的永不停止的追求导致了砷化镓(GaAs)晶体管的发展。尽管 GaAs 中电子的极高的迁移率在 20 世纪 50 年代就被认识到了，但是如何经济地生产有足够高纯度与足够低缺陷的 GaAs 的困难推迟了 GaAs 晶体管的有规模的商业化。这个情况一直延续到 20 世纪 60 年代后期。加州理工学院的 Carver Mead 成功地在 1965 年演示了第一个肖特基栅 GaAs FET(场效应管)，金属—半导体 FET(MESFET)到了 20 世纪 90 年代最终占据了手机的功率放大器(功放)市场的绝大部分。

蜂窝布局的手机最终在 1969 年年初以一种有限的形式出现在纽约市与华盛顿特区(美国首都)之间运行的火车上，用作为付费电话。这个 450MHz 的系统，尽管只是局限在这一单一交通路径，具备了蜂窝结构的定义性特征：频率复用与切换(handoff)[59]。几年之后，摩托罗拉申请了一个专利，它经常被引用为今天使用的蜂窝想法的第一个表述[60]。到了 1975

---

[54] 准确地说，是 1946 年 7 月 17 日。见(Telephone Service for st. Louis Vehicles)(提供给圣路易斯车辆的电话服务)，*Bell Laboratories Record*，1946 年 7 月。

[55] 但是，这个服务其实只提供半双工的操作：用户必须按下一个键来讲话，然后放开这个键来听。此外，所有的电话是由话务员来转接的，没有直接拨号的能力。

[56] 为了强调那里面是没有什么新的主意，贝尔自己就为他的"谐波电报"发明了一个初步形式的 FDMA，在这个通信系统中，一个共用的电报线为许多使用不同频率的用户分享。个别调谐的调谐叉保证了只有那些设定的接收者的电报才会得到响应。

[57] 与广泛相信的正好相反，为大多数微波炉所使用的 2.45GHz 频率并不相应于任何与水分子谐振的频率。更高的频率可以被更强地吸收但穿透深度差些。较低的频率则有更深的穿透深度，但加热起来不那么有效。存在着一个大的频率范围，其中穿透深度与加热速度取得合理的平衡。

[58] 见：H. Kragh，*Cosmology and Controversy*(宇宙学及其争议)，Princeton University Press，Princeten，NJ，1996.

[59] C. E. Paul，"Telephones aboard the Metroliner"(城际列车上的电话)，*Bell Laboratories Record*，March 1969.

[60] Martin Cooper 等，美国专利#3 906 166，1973 年 10 月 17 日申请，1975 年 9 月 16 日授权。贝尔实验室与摩托罗拉当时是比赛，哪一个能最先实现蜂窝这个概念。尽管贝尔实验室已经在理论方面工作了更长的时间，摩托罗拉被认为是第一个建造了一个系统级的原型，也是第一个完成了用手持移动电话进行蜂窝网的通话(如 Dan Gillmor 在 2003 年 3 月 29 日在圣荷西水星报报道的，但 Cooper 认为这个日子是 1973 年 4 月 3 日)。

年，贝尔系统公司最终接到 FCC（美国联邦通信委员会）的批准来提供试运行的服务，但是直到 1977 年它才得到允许来正式运行。试运行的服务最后是在 1978 年于伊利诺伊州的芝加哥市进行的，过渡到完全的服务则是在 1983 年 10 月 12 日最终发生。被称为 AMPS（advanced mobile phone service），这个模拟信号、基于 FM（调频）的系统工作在一个新近指定的 800MHz 左右的频带（这个频带是通过取回之前已给了 UHF 电视频带的高端频谱，而为它创立的）。就像 MTS 与 IMTS，AMPS 采用了频分多址（FDMA），在这个制式中多个用户可以通过给定的不同的频率同时通话。就像 IMTS，FDMA 也用了频分双工（frequency-division duplexing，FDD），使得一个用户一边听一边讲。这与普通电话的使用是一样的。我们可以回想起，在 FDD 程式中，是用不同的频率来分别进行发射与接收的。

当然，其他国家也在设计类似的系统。要逐一提及这些系统实在是太多了，但是可以特别提到的是 450MHz 的北欧（Nordic）移动电话系统（NMT-450，在 1981 年开始启用）。这是一个多国蜂窝系统——为芬兰、瑞典、丹麦与挪威提供服务。除了频率范围，它的特征与 AMPS 的十分类似。在十年内，第一代的蜂窝系统手机服务已经变得无处不在了[61]。这个没有惊喜地预计到的迅速成长推动了微波系统的成长。那些之前只是专门为军用与航空航天应用的发明，在消费类微波系统成长的推动下继续成为了目的在于降低成本与大批量生产的重要创新。

展望未来，人们可能会想象有一个在频谱上的类似的摩尔定律（Moore's law），驱使载波频率不断增高。然而，我们必须考虑到因为大气中的吸收效应而造成的信号衰减（见图 1.16），当频率达到几十个 GHz 时开始变得重要。在频率低于 40～50GHz 时，在海平面的大气吸收典型值是 1dB/km，但是遇到大雨，损耗会大大增加[62]。在 22～63GHz 这两个频

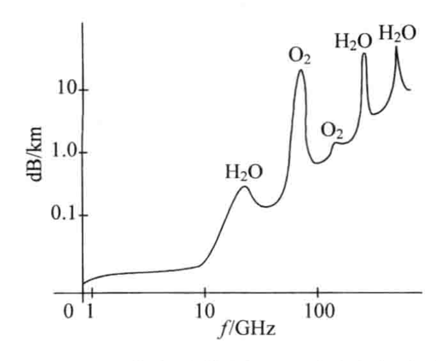


图 1.16　近似的在海平面上干燥空气中的大气衰减与频率的关系[63]

[61] 这样的增长使几乎所有的人都大为吃惊。在一个著名的（或许是挨人骂的?）由麦肯锡公司（McKinsey and Company）在 AT&T 要求下于大约 1982 年进行的研究中，预测整个美国手机市场到 2000 年会饱和在 900 000 高消费的用户。事实上，在 2000 年美国有超过 1 亿的用户，这样预测是错了 40 分贝以上。今天（指 2004 年本书出版时，译者注），每天在全世界卖出超过 1 百万部手机，而总的手机用户超过了 10 亿（是已在使用的 PC 数量的两倍）。根据麦肯锡研究的结论，AT&T 早早就出售了手机业务——只是当发现这个错误是过于大了而不能忽略时，又在 1993—1994 年付了 115 亿美元来重新进入这个市场。

[62] 这些值是在 Friis 路径损耗之上的。

[63] 根据 *Millimeter Wave Propagation: Spectrum Management Implications*（毫米波传播：频谱管理的影响），Federal Communications Commission, Bulletin（公报）no. 70, July 1997（1997 年 7 月）。

率附近(可能上下会差1GHz左右)有强的吸收峰。那个较低频率吸收峰是因为水分子,较高的那个则是因为氧分子。这个氧分子吸收的峰造成的路径损耗超过20dB/km,因此是十分严重的。如果想实现在短距离上的频率复用,则这个衰减很有用。这个性质在不少建议书上都被探究过,从皮可单元(picocell)到在60GHz下的其他短距离服务。这个大量的频谱资源使得实现高数据率成为可能,而差的传播特性恰恰可以被利用来强制高的频率复用性(就是说,这不是一个编程中的错误(bug)而是一个特点)。

最近,FCC将两个最低的氧分子吸收峰之间的频谱拨出来给商用的宽带无线应用,被称为"E-band(E频段)"。这个频段包括71～76GHz, 81～86GHz及92～95GHz,是用于高速、点对点无线局域网(WLAN)以及宽带互联网的接入。再重复一遍,这个相对高的大气吸收(在大的降雨中可能是10～20dB/km)被认为是优点来允许最大的频谱复用,而同时最小化干扰。另一个优点是容易用小尺寸的天线实现将短波长的发射形成紧密的波束。微波技术的发展几十年来一直是被军事、航空航天与无线电天文学的应用驱动的。但在过去的二十年中,有一股确定的朝消费者微波产品发展的趋势。以往的那些用手工做、小批量与不计成本的微波电子产品已经大多数让位给今天用完可以丢弃的手机与WLAN卡。工作在经典微波频段的产品的年销售量是惊人的。每天有超过一百三十万部手机被售出,看上去对通信器件的需求没有尽头。无线局域网、传呼机、卫星电视服务,GPS(全球定位系统),超宽带(UWB)连接,汽车防撞雷达及射频标签(RFID)仅仅构成了消费者微波应用的一个子集。而且我们一定不能忘记通过磁共振成像的核磁共振(这是二次大战后微波研究的后续成果,也受益于在战后雷达部件的过剩)。假若历史是未来的一个引导,我们几乎可以肯定我们还会为微波技术的发展方式感到惊讶。

1.5 结语

回顾移动与手持无线通信的历史,看起来好像500MHz～5GHz这个频段是过度地被使用了。这个看上去的钟爱并不是因有选择地有意挑明而造成的假象。频谱的这一段的的确确是十分受青睐的,这是有其内在的原因的。

首先,让我们考虑一下什么样的因素会限制无线应用在较低的载波频率下工作。一个简单的原因是那里频谱资源较少。然而,更重要的是天线不能做得太小(相对于波长而言),否则天线不会有效地工作。效率高的低频天线因而必然是长天线。对于移动或手持应用,我们必须选择足够高的频率,这样效率高的天线就不会太长。在500MHz,其自由空间的波长是60cm,四分之一波长天线大约是15cm长。这样的大小可以容易地放入一个手提单元。

当频率增加时,我们遇到了变坏的路径损耗。一个引发的因素是愈益增加的反射、折射与衍射的趋势,但是另一个因素能够从Friis公式预期到。将频率增加10倍,到5GHz,就增加了Friis路径损耗达20dB。在这些频率下,与生物组织的相互作用就不能忽略,因此简单地增加功率100倍来补偿不是一个解决方案。工作在更高的频率下是伴随着通信的不断减小的实际半径。

这样,我们看到了存在着大约10倍的频率区间,从500MHz～5GHz,这个区间将永远留为大面积移动无线通信的"甜区(sweet spot)"。因而,不像摩尔定律,有用的频谱不会随着时间指数地扩张。事实上,这个频谱本质上是固定不变的。这个真理解释了为什么电信

运营商在 20 世纪 90 年代末期发狂似地用几千亿美元竞标 3G(第三代)频谱(只是到后来发现所负的债务是如此之大,以至于许多运营商不得不"重新"进行协商)。无可置疑,人们一直在进行努力来使得这个有限频率范围的利用最大化,同时努力从别的利用率不那么高的无线应用频段(比如 UHF 电视广播)中取得频谱资源。

1.6 附录 A 其他一些无线系统的特征

不可能在这里将所有的在使用的无线服务与系统都一一列出,但是我们在这里提供一个简略的对其他有兴趣的服务的取样调查。(一个详细的美国频谱指定表可以从免费的网址下载 http://www.ntia.doc.gov/osmhome/allochrt.pdf。)这些频段的第一个是无需许可证(unlicensed)的 ISM(industrial-scientific-medical)频带,见表 1.1。微波炉、传感器、射频标签、一些无绳电话、WLAN 以及许多其他应用与服务使用这些频带。注意,这些频带位于先前指出的移动与手持无线通信的"甜区"。另一个无需许可证的频段最近在美国被批准。这个被称为"无需许可证的国家信息基础设施"(unlicensed national information infrastructure, UNII)的频带增加了 200MHz 到现在的 5GHz ISM 频段,同时也在其中一个频带里允许相当高的 EIRP,见表 1.2。

因为移动与蜂窝系统不是无线的仅有应用,表 1.3 给出了其他(广播)无线系统的一个简要的抽样信息。

表 1.1 ISM 频带划分与小结

参　数	900MHz	2.4GHz	5.8GHz
频率范围/MHz	902～928	2400～2483.5	5725～5850
总频带/MHz	26	83.5	125
最大功率/W	1	1	1
最大 EIRP[a]/W	4	4(200W 点对点)	200

[a] EIRP代表"等效各向同性的辐射功率",等于辐射功率与天线增益的乘积

表 1.2 UNII 频带划分与小结

参　数	Indoor	Low-power	UNII/ISM
频率范围/MHz	5150～5250	5250～5350	5725～5850
总频带/MHz	100	100	100
最大功率 mW	50	250	1000
最大 EIRP/mW	200(单元必须有集成天线)	1	200

表 1.3 一些广播系统的随机抽样信息

服务/系统	频率范围	频道间间隔
AM radia/kHz	535～1605	10
TV(ch. 2-4)/MHz	54～72	6
TV(ch. 5-6)/MHz	76～88	6
FM rodio/MHz	88.1～108.1	0.2
TV(ch. 7-13)/MHz	174～216	6
TV(ch. 14-69)/MHz	470～806	6

1.7 附录 B 谁是无线电的真正发明人

这个问题特意这么提出是有点挑衅的味道，其实这更像一个 Rorschach 测试（一种心理测试，Rorschach 是提出这种测试的瑞士心理学家的姓，译者注），而不纯粹是一段历史的较真（如果真是这么一个争论的话）。坦率地说，这只是一个借口让我们来回顾某些无线电先驱者的贡献，而不是一个真诚的努力来提供对这个问题的确定的（与肯定的）答案。

首先，用无线电或无线这个词，到底意味着什么？假若在最广义与宽松的意义上来诠释后面（即无线）这个词，那么或许不得不包括用烟火与旗语来传递信号这种形式，这个发明可是大大先于用电线来传送信号。你或许会争辩将定义扩展到如此地步明显是十分愚蠢的。但是如果我们然后将定义限制在用赫兹波的辐射来通信，那么就必须排除那些只是简单地将大气处理成为一个导体的技术。这一类技术包括许多拥有衷心支持者的发明家的贡献。我们可以基于这样的标准再细分，比如可商业化、实际可行性、调制方式，等等。缺乏关于无线电与发明这两个词到底意味着什么的共识是争议的核心，但我们不会去试图解决这件事。

一个先驱者是牙医 Mahlon Loomis，对他的热情支持者的数量超出了他所值得拥有的。他在 1872 年申请了一个有关无线电气通信的专利[64]。在一个使人想起富兰克林（Benjamin Franklin）电气实验的结构中，Loomis 提出了一个风筝系统来让一些电线飘起来。加在这些电线上的足够高的电压会让电信号通过大气而传递，然后接收器能用电流计来检测感应到的电流。据说，Loomis 在他家乡西弗吉尼亚州进行的实验是成功的，但是没有一个原始的证据来支持这个说法。而且基于现代知识的计算表明在任何一种情况下这个结论都是十分需要怀疑的[65]。Loomis 的支持者面临更为严重的困难，因为 William Henry Ward 就在 Loomis 三个月之前申请了一个想法基本上是一样的专利（而且采用了更为复杂的仪器），见图 1.17[66]。不言而明，通过大气可靠地导通足够的直流电流以在电流计中产生一个可以测量与不含混的响应基本上是没有希望的，而且不论是 Loomis，还是 Ward 都没有描述一个可以工作的无线电报的系统。

然后，David Edward Hughes 注意到他自造电话中的一个连接松的接触会对一定距离之外的其他的设备产生的电气扰动作出反应。经过了一些实验与改进，他在 1880 年 2 月 20 日给一个由皇家学会会长 Spottiswoode 为首的委员会作了他的发现的报告。这个演示包括了历史上第一个无线接收器。一个名叫 Stoke 的教授评论说，尽管看起来挺有趣，这个现象只不过是普通的磁场感应的动作，并不是麦克斯韦预见的验证。因为他的受尊敬的同

---

[64] 美国专利＃129 971，1872 年 7 月 30 日授权。这个一页的专利没有任何绘图。可以这么说，对于一个在这个领域内只具备一般能力的人，仅仅基于在专利里描述的信息是不可能将其发明实现的。Loomis 是一个行为十分古怪的人，从一些文章的题目看上去，他的有关写作好像是供娱乐性阅读的。

[65] 看起来有权威的有关成功测试的报道好像很多（在典型的说法中，从几个州来的参议员在测试时是在场的，这些测试是通过在相距 22km 的山顶间通信进行的。这个通信被演示与证实了。然而，除了 Loomis 自己提供的材料，我从来未能找到有关这些测试的信息。其他引用同样结果的报道明显地也没有更好的成功率来确定一个原始的来源，而是继续没有验证地重复这些说法）。

[66] 美国专利＃126 536，1872 年 4 月 30 日授权。

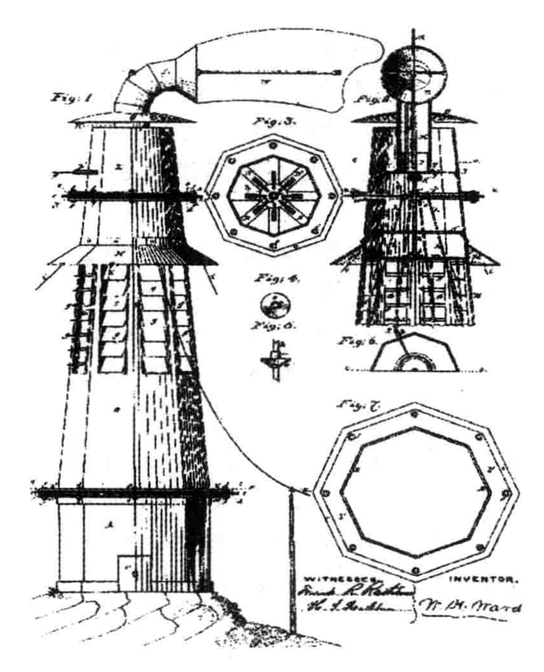


图 1.17 Ward 专利的第一页

事们如此强烈的判断，致使 Hughes 放弃了对无线电的进一步工作[67]。

在同一年（指 1880 年，译者注），贝尔（Alexander Graham Bell）发明了光线电话[68]，一个利用硒的光敏感性[69]来用于光学无线通信的装置。这个电话只能在大白天，且在发送与接收两端间没有光线阻挡物（line-of-sight）的条件下才能工作，光线电话从来就没有得到商业的应用，它充其量只是在无线的历史上的一个脚注而已。贝尔自己则认为这是一件十分重要的事，以至于他的四个专利是与这个光线电话有关的。

然而，基于大气传导的电报继续吸引着注意。Tufts 大学的 Amos E. Dolbear 教授在 1886 年申请了这种系统的另一个的专利，见图 1.18[70]。这个发明之所以值得注意主要是因为它直接承认了大气是一个被共享的媒介。为了确保多个用户平等地利用这个资源，Dolbear 建议给每个用户规定一个专门的时间段。这样，Dolbear 的专利是第一个为无线通信描述时分多址（time-division multiple access，TDMA）的。马可尼后来买下了对这个专利的使用权。

[67] Hughes 是如此之失望，以至于他甚至没有发表他的发现。这里的故事是根据 Ellison Hawks 写的 *Wireless Pioneers*（无线先驱者）（Methuen，London，1927）。而 Hawks 又是引用了 Hughes 在 1899 年发表的记录。

[68] A. G. Bell 与 S. Tainter，美国专利＃235 496，1880 年 12 月 14 日授予。

[69] 硒的这个性质也激起了在当时的用于电视的极其大量的专利。对硒的潜力的热情到了这种程度，以至于 *The Wireless & Electrical Cyclopedia*（无线与电机百科全书）（Catalog no. 20 of the Electro Importing Company，New York，1918）过分地预言“硒将在这个世纪里解决许多问题，它是一个从来没有被发现的最神奇的物质”。我觉得，你如果不在乎它的毒性，倒真的是那么回事。

[70] 美国专利＃350 299，1886 年 10 月 5 日授权。

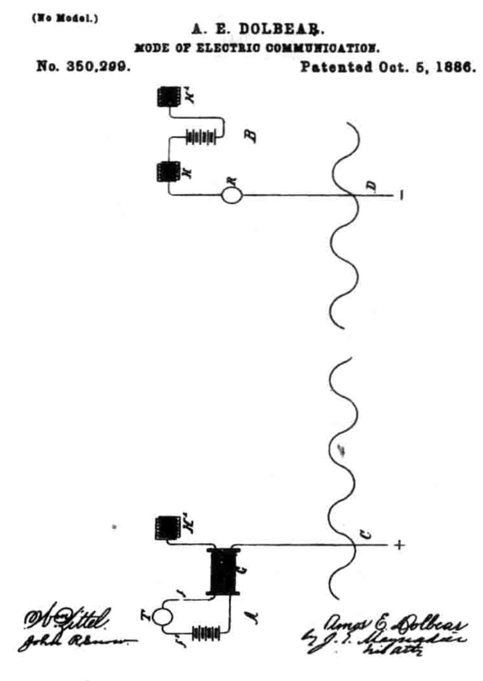


图 1.18　Dolbear 专利的第一页

我们当然也不能忘记赫兹(Heinrich Hertz)。他为他的 1886—1888 年的研究所构建的仪器与其后的无线通信先驱者用到的仪器其实没有什么差别。他对基本物理的倾心关注，而又过早地去世，是其他人因为无线通信的发明得到功劳的原因。

与赫兹一样，洛奇开始也没有全力倾注在用于通信的无线技术的应用上。例如，1894 年他在伦敦皇家学院(Royal Institution)一个会议上的演示(题名为"The Work of Hertz"(赫兹的工作)，这也是无线技术的相干器(coherer)的首次公开亮相)并没有包括有意识的信件发送与接收[71]。洛奇后来承认自己起初对无线通信缺乏兴趣是因为两个偏见：一个是有线通信是一个成熟与可靠的技术，很难想象无线通信技术能达到与有线通信一样的水平；另一个偏见或许是因为在同时知道得太多与太少。在证明了赫兹波与光波是一回事后，洛奇错误地得出这样一个结论，即无线通信会局限于光线不被遮挡时(line-of-sight)的通信，这样就限制了这个技术的商业应用前景。洛奇在这方面的偏见不仅是他一个人。大多数"专家"都持有同样的观点。尽管如此，他继续发展这个技术，并且他在连续波发生技术开发出来好多年之前就申请了用于无线通信的调谐天线与电路的专利(见图 1.19)。他是第一个提出共振(syntony)这个词来描述同步调谐的电路。读者可能已经注意到这个词没有被广泛使用。可怜的洛奇，几乎没有一件他做的事是到今天还被记得的。

洛奇发表的文章很多，且他的文章激发了波波夫(Alexander Popov)在俄国从事同样的研究[72]。波波夫在 1895 年 5 月 7 日将他的仪器演示给他的俄国物理与化学学会的同事们看。这个日子依然在俄国被庆祝为无线电日，尽管就像洛奇的先他一年前的演示，他的实验

[71] Aitken，*Syntony and Spark* (共振与火花)

[72] 波波夫的名字在其他场合也被拼成 *Aleksandr Popoff*(以及类似的变形)

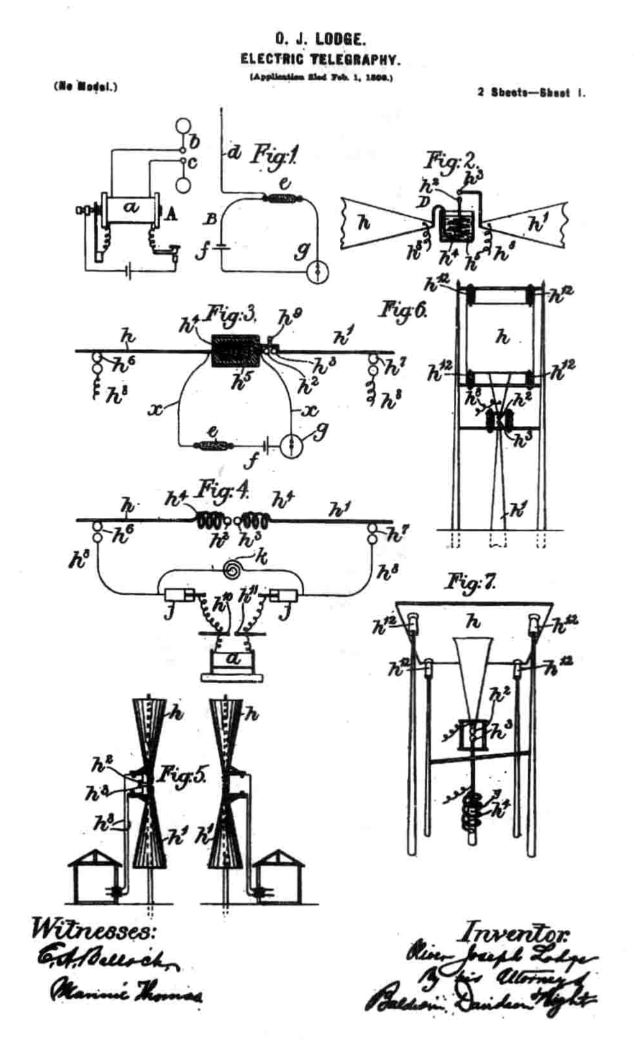


图 1.19　洛奇美国专利＃609 154 的第一页（1898 年 2 月 1 日申请，1898 年 8 月 10 日授权）

没有演变为实际的通信工具。

根据在这个事件三十年后写下的传说，波波夫在 1896 年 3 月 24 日演示了无线电报。这是一句"Heinrich Hertz"的话，通过大约 250m 距离发送与接收的。他在一年之后又演示了第一次船与岸之间的通信。对他的仪器的继续改进使得第一次无线协助的救援在 1899—1900 年间得以成功实施[73]。

不同于 Huges，Dolbear，赫兹，洛奇与波波夫，他们都是学术界精英的成员，马可尼（Guglielmo Marconi）却是一个社会圈子里的精英。他在 1894 年 12 月，读了 Righi 写的赫兹的悼念文章不久之后，开始奋力工作。到了 1895 年初期，他已经积累了足够的知识与设

[73] 这个日期的不一致性反映了确立回绕着波波夫贡献的一些问题中的一个。虽然来源不同，它们都具有明显相同的可信度，引述的日期从 1899—1901 年有关在芬兰湾的 *General Admiral Apraskin*（Apraskin 少将号）军舰的营救。很不幸的是这样一个重要的被认为是在 1896 年 3 月 24 日取得的成就在三十年内一直没有被记录。见 Charles Susskind，"Popov and the Beginning of Radiotelegraph（波波夫与无线电电报的起源）"，*Proc. IRE*，v. 50，October 1962.

备开始在他家庭的别墅(在 Griffone)里外进行实验。他一直记住这个技术的商业应用前景,因而在早期就申请了几个专利,并且在 1896 年 6 月 2 日得到了其中的第一个(英国专利 #12 039)。

根据有记载的证据,马可尼在波波夫之前就演示了称得上是真正的无线通信,尽管开始时只是面对一小群不属于学术或专业机构的人。马可尼与波波夫都没有用比洛奇在一年之前使用的在某些方面性能更先进的仪器。他们间的主要差别是从一个简单地演示无线电效应能够被发送这一事实转变到有目的地用无线电效应来进行通信。

这样看来,是否谁是无线电的发明人这个问题可以被简化到在马可尼与波波夫之间选一个?或者是马可尼与洛奇之间?或洛奇与波波夫之间?但是特斯拉(Tesla)又如何呢?

特斯拉?

Nikola Tesla(特斯拉)的同步电机的发明使得交流电的产生成为实际可行,并且引入了随之而来的世界电气化。特斯拉后来沉醉于无线传送具有工业应用所需要强度的功率的想法。基于他与低压下的气体打交道的经验,他知道气体是十分容易电离的,因此导电性变得十分好(这个行为是霓虹灯与日光灯的基础)。就像在他之前的 Loomis 与 Ward,特斯拉决定用大气作为导体。他进一步推断地球的上层大气,因为也是处于低气压状态,其导电性必然是很好的。特斯拉朝着开发要在地面与导电的上层大气之间形成导电通路所需要的极高压努力工作。特斯拉估计他需要几千万伏或更高来达到他的目标[74]。普通的交流升压变压器没有能力产生这样高的电压。著名的特斯拉线圈(现在是一个世纪来高中科学课程中不可缺少的物品)就是源于他的努力来建造一个实践中行得通的兆伏电压源。基于他对于谐振现象深入的理解,特斯拉线圈利用了谐振电路能提供的十分可观的电压提升。

特斯拉在这个系列专利中的第一个(见图 1.20)是美国专利 #645 576,1897 年 9 月9 日申请,1900 年 3 月 20 日授权。这个专利只提及电力能通过大气导通,但没有提到有关信息的传递[75]。

这个专利是在一个著名的 1943 年美国最高法院的裁决(320US 1,4 月 9-12 日辩论,6 月 21 日裁决)中引用的几个专利中的一个。这个裁决经常被当作是特斯拉从此确立了作为无线电发明者的地位。这个案件的背景是:美国马可尼无线电报公司在第一次世界大战后不久控告美国政府在几个无线通信专利中是侵权行为。法院的裁决十分清楚地说马可尼有关四谐振电路系统的专利是无效的,因为已经有先前的技术存在。在三个其他控告美国政府侵权的专利中,一个被裁决没有侵权,另一个是起诉不成立,而第三个则被认为起诉是成立的并判美国政府侵权。结果是美国政府被判支付一个象征性的大约 43 000 美元。这个 1943 年的裁决通过引述洛奇、特斯拉与 John Stone 的先前发明来否定四电路的专利(这个专利在英国已经作为专利 #7777 生效),终结了这个争议。这个裁决就这样明确地宣布

---

[74] 后来,他开始在纽约州的长岛建造一个巨大的塔来无线传输功率。为著名的 Gilded Age(镀金年代,指美国内战后 1877—1917 年期间美国经济与人口快速发展的年代,译者注)建筑师 Stanford White(他的被暗杀被记录在 *Ragtime*)所设计,这个以装有令人印象深刻的紫外线(UV)灯阵列为特征的 Wardenclyffe 塔,明显是要通过 UV 离化来帮助产生一个导电性更好的通路。因为没有足够的资金,这个塔从来就没有被完成。塔的部件最终是当作废料被出售,而建筑剩下的部分被推平了。

[75] 他之后的专利的确讨论了信息的传递,但他的阐述有意识地排除了赫兹波的使用。他完全是沉醉在用地球作为一个导体,大气为另一个导体来输送电力。

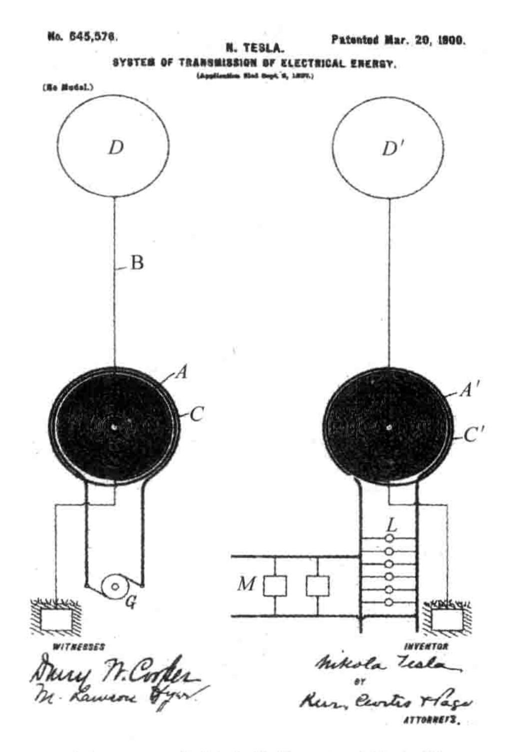


图 1.20　特斯拉的第一个无线专利

马可尼不是这个电路的发明人，但它也并没有说马可尼没有发明无线电。裁决的确指出了四谐振电路系统使得第一个实践上行得通的基于火花的无线通信得以实现(这个四谐振电路系统与连续波系统基本无关)。但是，法院并没有作出一个跳跃式的结论说洛奇、特斯拉或 Stone 其中任一人因而是无线电的发明者[76]。这个经常被引用的裁决因此实际上没有对发明权作出明确的结论，只是作了一些负面的说明。人们可能会猜测，最高法院有意识地避免作出影响更为大范围的声明是因为完全认识到发明与无线电这两个词的含义都不唯一。

我们能够确定地说的只是，在这些早期的先驱者中，马可尼是第一个相信无线通信不只是一个知识界的练习。他固然没有在电路领域中创新(并且我们可以不是偶然地说，他是改造性地采用了别人的发明)，但他的远见与将无线通信作为一个重要的商业行为的决断是得到了回报。因为他很快地作出了十分关键与重要的发现，即无线不只是局限于射线不能被挡住的通信，这证明了专家的观点是错的。马可尼几乎是单独地相信无线电是一个重要的商业运作而使其成为一个重要的技术。

如此说来，到底是谁发明了无线电？就像我们已经说过了，这取决于你对发明与无线电这两个词是如何定义的。假若你是指第一个想到用某种形式的电信号来进行无线通信，那么 Ward 会是一个候选人。假若你是指第一个建造了基本的用波的无线技术仪器，那么赫兹比谁都合适(并且因为光就是一种形式的电磁波，我们会不得不包括贝尔与他的光电话)。假若你是指第一个来用赫兹波有目的地传送信息，那么或者是波波夫，或马可尼会是一个靠

[76] 欢迎读者来独立地验证本书作者的判断。整个案件可以在线查到：http://www.uscaselaw.com/us/320/1.html。

得住的选择(然后,又要轮到贝尔了,他是最早就将光电话专门用来进行通信)。假若你是指第一个看到调谐对无线通信有价值的,那么洛奇以及或许是特斯拉都是发明者。但有证据表明洛奇会是一个更强的候选人。

面对这一系列的可以信赖的选择,认为这个或那个人(发明了无线电)就不奇怪了,因为其中是有国家或其他感情因素在起作用的,而不是纯粹基于技术上的判断。类似于这种情形的例子并不少,致使肯尼迪(John F. Kennedy)总统作出了"成功总是有许多父亲的"这样的观察。无线电技术当然是一个巨大的成功,因此遇到这么多人宣称为其父亲就一点都不奇怪了。

第2章 射频与微波电路导论

2.1 一些定义

这一章的标题会引起一、两个问题：射频(RF)与微波(microwave)的确切定义是什么？我们在前面一章中用到了这些术语，但有意识地不提供一个定量的定义。某些教科书用绝对的频率来进行区分(比如说，“微波是1GHz以上的信号”)。然而，这些词的意义是随着时间发生变化的，表明基于绝对频率值的划分是缺乏基本的根据。的确，就工程实践与设计的直觉而言，下面的做法会更有价值，即通过将一个电路元件的物理尺度与在其中传播的信号波长的比较来进行分类。

就如我们很快可以明白的，当电路的物理尺度相比于所涉及的波长非常小时，我们是处在通常的电路理论范畴。将这个情形称为准静态、集总，或低频范畴，而不管实际的频率值。这个尺度的不相称性大大简化了麦克斯韦尔方程，使得我们可以采用电感、电容与基尔霍夫电流与电压定律这些熟悉的概念。

在另一方面，如果物理尺度比涉及的波长要大得多，那么我们说系统是运作在经典的光学领域——不论所涉及的信号是否相应于可见光。用来处理能量的器件结构现在是镜子、偏光器、透镜及衍射光栅等。就如在准静态情形那样，尺度的不相称性也大大地简化了麦克斯韦尔方程。

假若电路的物理尺度与所涉及的波长大致相当，那么就有了称之为微波或分布范畴。在这个处于中间范围的区域，过多地化简麦克斯韦尔方程(如果真的可以这样做的话)，使得我们取得设计的内在理解十分复杂[①]。比如说，可以在谈及反射的同时来讨论电感、电容。也可能出现这样的情形，试图确定单个的电感与电容或者是不恰当的，或者是根本不可能的，因为能量会同时储存在处于同一空间区域的电场与磁场中。在微波领域，我们需要面对传输线行为，而且能量的辐射(不管是企求的或是不希望的)会潜在地变得十分严重。这本书的主要关注点是将集总范畴与分布领域桥接起来。

这样，用于指导系统设计的技术及直觉最好是基于这些归一化后的波长范畴划分。作为一个结果，会有许多频率上的交叠(比如说，我们可以做出完美又性能很好的工作在GHz频率范围内的菲涅尔透镜，如果我们愿意建造相当大的结构的话)。以上的观点解释了为什么一个基于任意频率界限的分类系统并不像我们所希望的那样有用。

作为对这个标题的最后一个评论，在这本书里，我们通常忽略光学范畴。然而，读者不应该从这个忽略上得出这样一个结论，即相对于通常被称为射频电路设计而言，光学技术是没有用或无关的。这主要是因为我们的书必须在某个地方收笔。

---

① 光纤的截面尺度与波长可相比拟，因此对它们的分析落入这个中间区域，而不是通常的光学领域——这就是我们要区分经典光学与这个中间范畴的原因。

2.2　常规的频段划分

表 2.1 列出了频段划分与它们通常的(但不是普适的)标识,这看起来与前文关于引用任意频率边界在物理上不太正确的说法不太吻合。有关这些频段的精确界定(尤其是对那些低于 VLF 与高于 UHF 的频段)并不是所有的来源是一致的,因此最好是在引用这些频段时,附加说明实际的频率值。

表 2.1　无线电频段的标定

频　　段	频 率 范 围	波 长 范 围
极低频(ELF)	<30Hz	>10 000km
超低频(SLF)	30～300Hz	10 000～1000km
特低频(ULF)	300～3kHz	1000～100km
很低频(VLF)	3～30kHz	100～10km
低频(LF)	30～300kHz	10～1km
中频(MF)	300kHz～3MHz	1km～100m
高频(HF)	3～30MHz	100～10m
很高频(VHF)	30～300MHz	10～1m
特高频(UHF)	300MHz～3GHz	1m～10cm
超高频(SHF)	3～30GHz	10～1cm
极高频(EHF)	30～300GHz	1cm～1mm
荒唐的高频(LHF)	>300GHz	<1mm

在将波长与频率联系起来时,记住频率(单位: Hz)与波长(单位: m)的乘积等于光速(非常接近于 3×10^8m/s)。因此,一个 1MHz 的信号在自由空间中的波长几乎精确地等于 300m; 一个 1GHz 的信号波长是 300mm。下面的规则是十分有用的: 当频率的单位是 GHz 时,乘上以 mm 为单位的波长,其积大约是 300。

另一个可替代的分类系统可追溯其起源到第二次世界大战期间的雷达研究工作。这是基于最初时随机选择的字母,目的是混淆敌人。这种缺乏标准化的做法的确是成功地让每个人(不只是敌人,译者注)都搞混了[②]。与这些字母相关联的频率范围随着时间的推移已经发生了变化,而且它们从一个国家到另一个国家(甚至在同一国家内)都是不同的,在不同的公司间也是如此。因为这些原因,基于字母的频段标识或许最好被认为是过时了,而应该避免使用(或至少是辅之于确切的频率值,就像先前提到的标识那样)。不管怎么说,这些标识依然在用,因此我们这里给出这些频段的一张由 IEEE 521—1984 定义的表(表 2.2)。这是本书作者知道的唯一的国际标准。频段 Ku 与 Ka 的标识分别来自“under K”与“above K”。

② 战时对保密的要求是如此之高,以至于在盟军(Allied)雷达设计圈子内甚至不同研究组之间所使用的术语也是非标准化的。我们依然生活在这个混淆局面的继承中,也许将来也是如此。

表 2.2　微波频段定义与标识(IEEE 521—1984)

频　　段	频率范围/GHz	频　　段	频率范围/GHz
L	1.0～2.0	K	18～27
S	2.0～4.0	Ka	27～46
C	4.0～8.0	V	40～75
X	8～12	W	75～110
Ku	12～18		

读者可能遇到的字母标识的其他系统是波导频段及那些起源于 NASA(美国国家航空航天局)，惠普公司，Sperry，Motorola(摩托罗拉)，Narda，Raytheon，及其他一些公司与组织。这些标识都会有一小点不同，它们可能包括那些有额外字母的，或缺字母的标识频段。当调研文献时，要注意这一点。

普通的调幅(AM)无线电信号是在 MF 频段，而调频(FM)无线电工作在 VHF 频段。其他较低的电视频道(2～13，分布在 3 个不相连通的，跨越调频无线电频道的频段中)也在 VHF 频段内。电视频道 14～69 用频谱的 UHF 部分，同样在这部分频谱的包括所有目前的蜂窝手机通信系统、全球定位系统(GPS，频率是 1.575GHz)、微波炉(2.45GHz)，以及无绳电话(900MHz～2.4GHz)。警用的雷达随着时间的推进，工作频率不断升高，起先是 10GHz 的 X 频段系统，然后演变为 K，再到 Ka 频段(以及甚至是基于激光的光学系统)。这样做是为了始终比消费者能具有的且不断得到改进的雷达检测技术更胜一筹。

在上述频率利用单子中，看起来 UHF 频段的占用率相当高，这并不是我们有意地选择而造成的人为假象。与微处理器的世界不一样，那里是摩尔定律能有规律地提供指数式增长的计算资源，UHF 频段对于千米尺度的地面无线通信来讲是一个“甜区”(sweet spot)，而且永远会是如此。就像我们在第 21 章中将会看到的，为了与一个潜在的辐射器有效地耦合，通常要求天线的尺寸不比波长小多少(即，至少是 1/5～1/10 的波长)。因为无线通信通常也意味着移动通信，人们必须经常考虑天线的大小要与便携要求相符。假若我们随意地假设 10cm 为可以容忍的天线长度的上限，那么对天线有效辐射的要求就迫使我们不得不选择不大于 0.5～1.0m 的波长(就是说频率不低于 300～600MHz，因而一些早期的蜂窝电话系统工作在 450MHz)。

随着频率的增加，构造效率高的天线变得愈发容易。然而，传播就迅速变坏了，因为无线电波的特性越来越像光波了，会经历更大的衰减与衍射。更糟的是，生物吸收效应变得更严重，这样就限制了我们能够使用的用于克服路径损耗的辐射功率(要记住：一个典型的微波炉的工作频率就是 2.45GHz)。结果是，要采用几个 GHz 以上的频率来实现广域地面网络是十分困难的。最近发展起来的无线局域网(如 IEEE 802.11a)所用的 5GHz 频段或许就代表了实际的频率上限(也可能会在其一个倍频之内)。这个有用的，同时可被使用的频谱(大约是在从 500MHz～5GHz 的十倍频率间隔内)解释了为什么在竞标这些频段时付出的高昂代价，也是作为我们这本书讨论的焦点的原因：频率直到(大约为)实际可行的地面无线通信频段的分立射频电路的设计、构建与测量。

然而，如果我们给读者留下了这样一个印象，即不在本书讨论范围内的电路与系统是不值得考虑的，那就错了。对于那些不需要工作在大范围的地面区域，或者增大了的大气吸收反而是个优点的应用，或许知道下面这些事实是有用的：水蒸气在 22GHz 左右吸收最强，

而氧分子是在 63GHz(见图 2.1)。经常利用这些吸收峰进行保密通信,特别是在卫星之间,因为这样地面站的窃听装置除了噪声外什么也检测不到。同样地,汽车防撞雷达系统的工作频率要足够高(比如 77GHz)以提供所需要的空间分辨率,而同时利用高的自由空间衰减以减少汽车之间的干扰。有关"皮可网"(piconet,即微微网)的建议工作在类似的频段中(比如 60GHz),也是为了寻求利用在这些频率下信号在空间中的低的交叉互扰。采用天线阵列来产生高功率密度的窄波束可以帮助减弱在这些更高的频率下的路径损耗。对于这些频段的潜力的关注在美国联邦通信委员会颁布的将 V 与 W 波段一起称为"E 波段"可见一斑。

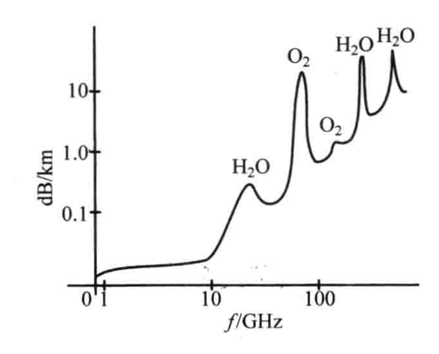


图 2.1　在海平面高度上干燥空气中的近似大气衰减与频率的关系[③]

有了这些材料作为背景,我们现在转入到本书的主题,首先来比较仔细地看一下集总电路与分布电路到底指的是什么。

2.3　集总电路与分布电路

正如我们已经说明过的,存在着两个重要的工作频率范畴,它们是根据电路元件是否可以处理成是"集总"的还是"分布"的来区分的。在这两个范畴间的模糊的边界是根据电路的物理尺度与我们所感兴趣的最短波长之间的比率来确定的。在足够高的频率下,电路元件的尺寸变成可与波长相比拟,因此我们不能想当然地采用从集总参数电路理论中得出来的直觉的知识。于是必须将导线看成传输线(正如它们本来就是那样的),基尔霍夫"定律"不再普遍适用,并且 R、L 和 C 的确定也不再那么明显(或者甚至是不可能的)。

为了在集总和分布领域之间划一条恰当的界线,我们需要(简要地)回顾一下麦克斯韦尔方程。

麦克斯韦尔和基尔霍夫

许多学生(遗憾的是也包括许多正在实践的工程师)忘记了基尔霍夫电压和电流"定律"是只在集总参数的范畴(这个概念我们一直在用,但却尚未定义)内成立的近似。如果我们假设准静态行为,因而消除了会产生波动方程的耦合项,那么基尔霍夫定律可以从麦克斯韦

③　根据 *Millimeter Wave Propagation: Spectrum Management Implications*(毫米波传播:频谱管理潜在影响),Federal Communications Commission(美国联邦通信委员会),公报(Bulletin)no. 70,1997 年 7 月。

尔方程中推导出来。为了理解这里所指的含义，让我们回顾一下微分形式的麦克斯韦尔方程（在自由空间）：

$$\nabla \cdot \mu_0 \boldsymbol{H} = 0 \tag{1}$$

$$\nabla \cdot \varepsilon_0 \boldsymbol{E} = \rho \tag{2}$$

$$\nabla \times \boldsymbol{H} = \boldsymbol{J} + \varepsilon_0 \frac{\partial \boldsymbol{E}}{\partial t} \tag{3}$$

$$\nabla \times \boldsymbol{E} = -\mu_0 \frac{\partial \boldsymbol{H}}{\partial t} \tag{4}$$

第一个方程说明不存在纯粹的磁荷（也就是不存在任何磁单极）。如果真的有纯磁荷，那它就会引起磁场的散度。我们以下将完全不会用到这个方程。

第二个方程（高斯定律）说明存在着纯电荷，并且说明了电荷是电场散度的来源。我们也不会真正去用这个方程。

第三个方程（安培定律，麦克斯韦尔对它作了著名的修改）说明无论是“通常”的电流还是电场的时间变化率都会有同样的效应：产生磁场。涉及电场导数的那一项是著名的位移（电容性）电流项，是麦克斯韦尔“无中生有”地变出它来以形成波动方程的。

最后，第四个方程（法拉第定律）说明一个变化的磁场会引起电场的旋度。

波动行为的起源，从根本上来说是因为在后两个方程中的耦合项：$\boldsymbol{E}$ 的变化引起了 $\boldsymbol{H}$ 的变化，$\boldsymbol{H}$ 的变化又引起了 $\boldsymbol{E}$ 的变化，如此往而复始。如果我们可以使 μ_0 或 ε_0 为零，那么耦合项就会消失，波动方程也得不到，于是电路就可以基于一种准静态（或甚至是静态）来加以分析。

作为一个具体例子，设 μ_0 为零就会使电场的旋度为零，这使 $\boldsymbol{E}$ 可以表示成电位梯度（带上一个负号）。于是这相当于以下的结论，即电场 $\boldsymbol{E}$ 沿任何封闭路径的线积分（这就是电压）为零：

$$V = \oint \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{l} = \oint (-\nabla \phi) \cdot \mathrm{d}\boldsymbol{l} = 0 \tag{5}$$

这里的最后一个方程只不过是基尔霍夫电压定律（KVL）的场论形式的表达式。

为了推导 KCL（基尔霍夫电流定律），我们以相同的方式进行但现在设 $\varepsilon_0 = 0$。于是 $\boldsymbol{H}$ 的旋度只取决于电流密度 $\boldsymbol{J}$。这使我们可以写出

$$\nabla \cdot \boldsymbol{J} = \nabla \cdot (\nabla \times \boldsymbol{H}) = 0 \tag{6}$$

也就是 $\boldsymbol{J}$ 的散度同样为零。没有散度意味着在节点上没有任何净电流产生（或丢失）。

自然，μ_0 和 ε_0 实际上都不会为零。为了说明前面的讨论并不是完全没有意义的，回想一下光的速度可以表示成④

$$c = 1/\sqrt{\mu_0 \varepsilon_0} \tag{7}$$

因此设 μ_0 或 ε_0 为零相当于使光的速度为无穷大。因此 KCL 和 KVL 实际上是假设了无穷快传播的结果；只要电路元件的物理尺寸与波长相比时较小因而光速的有限性不十分明

④ 应用中世纪哲学家 Occam 的“刀片”逻辑推理法（即尽量用最少数目的假设进行推理，译者注）的“鸭子确认”版本（“如果它走起来像鸭而且叫起来也像鸭，那么它必定就是鸭”），麦克斯韦尔证明了光和电磁波是一回事。毕竟，假若它以光的速度传播并且像光一样反射，所以它必定就是光。大多数人都会同意麦克斯韦尔方程的推导出来代表着 19 世纪知识上的至高无上的成就。

显，那么我们就可以认为它们(KCL 和 KVL)非常合理地成立，亦即当

$$l \ll \lambda \tag{8}$$

式中 l 是电路元件的长度而 λ 是我们感兴趣的最短波长。

为了对这一约束条件在数值上有个感觉，考虑一个最长尺度为 1cm 的电路元件。如果我们随意地定一个“小很多”是意味着“至少小 10 倍”，那么当信号最高频率对应的波长约大于 10cm 时这个元件可以被处理成是集总的。在真空中，这一波长相应于约 3GHz 的频率。在典型的电路板的材料上，这个频率要低 2 倍或更多一点。这个计算揭示了一个简单的真理：在几个 GHz 的频率范围中，人体大小的物体是不能按集总电路元件处理的。

总结一下，在集总电路理论(这里 KVL 和 KCL 成立，并且我们可以确定单个的 R，L，C)以及分布系统(这里 KVL/KCL 不成立，并且 R，L 和 C 不能是局域的)之间的界限取决于电路元件尺寸和所感兴趣的最短波长之间的相对大小。如果电路元件(就这里而言，互连线当然是一个电路元件)与波长相比时非常短，那么我们可以采用传统的集总概念而不会引起多少误差。如果不是这样，那么采用集总的概念是不合适的。在这本书中，我们将考虑其尺度与信号波长可以比拟的电路，因此，必须考虑分布效应。

2.4　集总与分布域的联系

我们现在转来考虑把在集总范畴中建立的设计直觉延伸到分布范畴中的问题。这样做的目的不只是为了教育，因为正如我们将要看到的，在这一过程中将产生对设计极为有价值的深入理解。一个重要的例子就是：用时延换取带宽，而不是用增益。

互连线是这样一个系统的例子，其在较低频率下可以被成功地作为一个简单的 RC 线处理。基于这一想法，减小 RC“寄生”参数以提高带宽已成为电路和系统设计者(特别是 IC 设计者)的主要关注点。遗憾的是把寄生参数减小到某个最小量之下在实际上是不可能的。因此从集总电路设计中得来的直觉会让我们误以为带宽是受限于这些实际上是不可能减小的寄生参数。幸运的是，把互连线恰当地处理为传输线，而不是有限的集总 RC 网络，表明这并不是这么回事。我们发现仍然可以以极宽的带宽来传送信号，只要认识到(实际上是利用)互连线的这一确实存在的分布性质。通过利用而不是去克服分布电容(和电感)可以由此实现延时和带宽之间的分离。这一新的观察结果是极有价值的，并且它同时适用于无源网络和有源网络。除了使我们能深入体会传输线现象之外，这一理解也将使我们得到某些放大器的拓扑形式，它们可以大大缓解表征低阶集总系统的增益-带宽间的折中(tradeoff)。

由于这些重要理由，我们现在通过延伸集总电路的分析来进行分布系统研究。

2.5　级联重复结构的策动点阻抗

我们从研究均匀的、重复结构的策动点阻抗开始。虽然要注意的是某些非均匀结构(例如按指数形式变化的渐变传输线[⑤])具有极为有用的特性，但我们把目前的讨论仅局限在考

⑤　对于不了解此点而又有兴趣的读者，我们可以指出，按指数形式变化的渐变传输线可以取得宽带阻抗匹配而不是由 1/4 波变换器提供的窄带匹配。变换比可以通过选择渐变变化比例常数来控制。

虑均匀的结构中。

具体来说，考虑如图 2.2 所示的无穷级梯型网络。尽管图中采用了电阻符号，但它们代表了随意的阻抗。

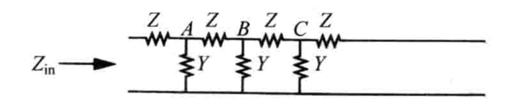


图 2.2　阶梯状网络

为了不用求一个无穷级数的和就能得出这一网络的策动点阻抗，注意到在节点 C 向右看的阻抗与在节点 B 向右看的阻抗是相同的，也与在节点 A 向右看的阻抗相同[⑥]。因此我们可以写出

$$Z_{in} = Z + [(1/Y) || Z_{in}] \tag{9}$$

上式可展开为

$$Z_{in} = Z + \frac{Z_{in}/Y}{1/Y + Z_{in}} \Rightarrow (Z_{in} - Z)\left(\frac{1}{Y} + Z_{in}\right) = \frac{Z_{in}}{Y} \tag{10}$$

求解 Z_{in} 得到

$$Z_{in} = \frac{Z \pm \sqrt{Z^2 + 4(Z/Y)}}{2} = \frac{Z}{2}\left[1 \pm \sqrt{1 + \frac{4}{ZY}}\right] \tag{11}$$

在 $Z=1/Y=R$ 的特定情况下

$$Z_{in} = \left(\frac{1+\sqrt{5}}{2}\right)R \approx 1.618R \tag{12}$$

这一 Z_{in} 对 R 的比率（或这个比率的倒数）称为黄金比率（或黄金平均值，或黄金分割）。它出现在各种各样的情形中，包括像希腊几何学家的美学、文艺复兴时代的艺术与建筑，以及对几个有趣的（但大部分是无用的）网络问题的答案。

理想传输线看作为无穷长的阶梯网络

让我们现在考虑在极限情况下输入阻抗的更一般的情形，即 $|ZY| \ll 1$，并且我们继续不允许 Z_{in} 取负值。在这种情况下，可以把结果简化为

$$Z_{in} \approx \sqrt{Z/Y} \tag{13}$$

我们看到，如果 Z/Y 碰巧与频率无关，那么输入阻抗也将与频率无关[⑦]。这类网络的一个重要例子是对一根理想传输线的模型。在一根无损耗传输线的情形中，$Z = sL$ 及 $Y = sC$，式中 L 和 C 代表微分（在数学意义上的）电路元件。因此一根理想的、无损的、无穷长传输线的输入阻抗（称为特征阻抗 Z_0）为

$$Z_{in} \approx \sqrt{Z/Y} = \sqrt{sL/sC} = \sqrt{L/C} \tag{14}$$

由于 Y（为一个无穷小电容的导纳）在微分元件的长度趋于零时趋近零，而微分电感元件的电抗也同时趋近于零，所以比例 $1/YZ$ 趋近于无穷大，因而满足为使我们上面的推导成

⑥　这是分析这种结构的一个极为有用的技巧，但令人吃惊的是有很大一部分工程师从来没有听说过这个方法，也或许是没能记住。不论怎么说，它相比于用更直接的方法无疑能节省极大量的工作，因为直接方法需要求各种无穷级数的和。

⑦　具有这一特性的阶梯状网络称为“常数 k”传输线，因为可以写成 $Z/Y = k^2$，k 为一个常数。

立所必需的不等式。这一结果——即对于一条无穷长传输线我们得到了纯实数的输入阻抗——应当是我们熟悉的结果，但也许得到这一结果的这个特殊的方法我们却并不可能熟悉。

一个经常会问的问题是，我们输入到一条传输线上的能量到哪里去了。如果阻抗是纯实数，那么这条传输线的特性应当像一个电阻，因此它应该将像一个电阻那样消耗能量。但是这条传输线是由纯电抗(因而是无损耗)元件构成的，所以在这里似乎出现了自相矛盾。

对这一问题的解答是如果这条传输线真的是无穷长的话，那么最终能量并没有变为热。能量只是沿传输线永远地传下去，所以能量损失给了外部世界，就仿佛它把一个电阻和它的周围加热一样；这条传输线的行为像是一个能量的黑洞。

2.6 传输线更详细的讨论

前面一节研究了一条无损无穷长传输线的阻抗行为。我们现在把我们对于特征阻抗 Z_0 的推导扩展到包括有损耗在内的情形。我们同时引入另外一个描述参数，即传播常数 γ。

2.6.1 有损传输线的集总参数模型

为了推导一条有损传输线的有关参数，考虑传输线的长度为 dz 的无穷短的一段线，如图 2.3 所示。这里元件 L，R，C 及 G 全都是每单位长度的量，并且只是前面所考虑过的比较一般情形中的一个特例而已。

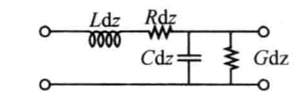


图 2.3 无限短传输线段的集总参数 RLC 模型

电感代表了存放在这条线周围磁场中的能量，而串联电阻代表了所有通常导体都存在的不可避免的能量损失(如由于趋肤效应，见 6.2 节)。并联电容模拟了存放在这条线周围电场中的能量，而并联电导则代表了由于如通常的欧姆漏电造成的损失以及在传输线电介质材料中的损耗。

2.6.2 有损传输线的特征阻抗

为了计算一条有损线的阻抗，我们按照与 2.5 节完全类似的方法：

$$Z_0 = Z\mathrm{d}z + [(1/Y\mathrm{d}z) \| Z_0] = Z\mathrm{d}z + \frac{Z_0}{1 + (Y\mathrm{d}z)Z_0} \tag{15}$$

因为我们将考虑当 dz 趋近于零时这一表达式的极限情形，所以可以利用 $1/(1+x)$ 的一阶二项式展开式：

$$\begin{aligned} Z_0 &= Z\mathrm{d}z + \frac{Z_0}{1 + (Y\mathrm{d}z)Z_0} \\ &\approx Z\mathrm{d}z + Z_0[1 - (Y\mathrm{d}z)Z_0] = Z_0 + \mathrm{d}z(Z - YZ_0^2) \end{aligned} \tag{16}$$

为使上式对任意的 dz 都成立，我们可以看到最后一项括弧中的量必须等于零。因此特征阻抗为

$$Z_0 = \sqrt{\frac{Z}{Y}} = \sqrt{\frac{R + \mathrm{j}\omega L}{G + \mathrm{j}\omega C}} \tag{17}$$

如果与电阻有关的项可以忽略(或者如果 RC 恰好等于 GL)，那么 Z_0 的公式就简化成我们前面推导的结果：

$$Z_0 = \sqrt{L/C} \tag{18}$$

因为在足够高的频率时，阻抗趋近于$\sqrt{L/C}$而与R或G无关，所以Z_0常常也被称为瞬态阻抗、冲激阻抗或脉冲阻抗。这个阻抗就是用来衡量同轴电缆品质的阻抗。根据公式(17)，这种电缆在直流下的阻抗在一般情况下与Z_0是十分不同的，这就解释了为什么“50Ω”电缆的阻值如果用欧姆计(这是直流测量，译者注)来测量，从来就得不到哪怕与50Ω这个数接近的值。

2.6.3 传播常数

除了特征阻抗外，另一个重要的描述参数是传播常数，通常用γ来表示。虽然特征阻抗告诉了我们在一条无穷长线上任何一点处电压对电流的比，而传播常数却能使我们了解有关在这样一条传输线上任何两点之间的电压(或电流)的比。也就是说γ定量描述了这条线的衰减特性。

考虑在一给定线段的两个端口上的电压。这两个电压的比很容易从通常的电压分压关系中计算出来：

$$V_{n+1} = V_n \left\{ \frac{Z_0 \parallel \frac{1}{Y\mathrm{d}z}}{Z\mathrm{d}z + \left[Z_0 \parallel \frac{1}{Y\mathrm{d}z} \right]} \right\} \tag{19}$$

于是(译者注：n沿传播方向增加)

$$\frac{V_{n+1}}{V_n} = \frac{Z_0 \parallel \frac{1}{Y\mathrm{d}z}}{Z\mathrm{d}z + \left[Z_0 \parallel \frac{1}{Y\mathrm{d}z} \right]} = \frac{Z_0}{Z_0 ZY(\mathrm{d}z)^2 + Z_0 + Z\mathrm{d}z} \tag{20}$$

由于我们将在$\mathrm{d}z$非常小的极限情况下运用这一表达式，我们可以忽略与$(\mathrm{d}z)^2$成正比的项，并且再次利用$1/(1+x)$的二项展开式，但仅保留$\mathrm{d}z$的一阶关系(记住，我们是工程师，整个世界对我们来说只是一阶的!)。这样就得到

$$\frac{V_{n+1}}{V_n} \approx \frac{Z_0}{Z_0 + Z\mathrm{d}z} = \frac{1}{1 + (Z/Z_0)\mathrm{d}z} \approx 1 - \frac{Z}{Z_0}\mathrm{d}z = 1 - \sqrt{ZY}\,\mathrm{d}z \tag{21}$$

尽管我们的推导不十分严密，但这些近似表达式的净误差在$\mathrm{d}z$为零的极限情况下确实收敛至零。

让我们把上面的方程重写为差分方程

$$V_{n+1} = V_n(1 - \sqrt{ZY}\,\mathrm{d}z) \Rightarrow \frac{V_{n+1} - V_n}{\mathrm{d}z} = -\sqrt{ZY}\,V_n \tag{22}$$

在$\mathrm{d}z$趋于零的极限情况下，这一差分方程变成了微分方程

$$\frac{\mathrm{d}V}{\mathrm{d}z} = -\sqrt{ZY}\,V \tag{23}$$

我们对这个一阶微分方程的解应当很熟悉

$$V(z) = V_0 \mathrm{e}^{-\sqrt{ZY}z} \tag{24}$$

亦即，在任何位置z处的电压只是电压V_0(在$z=0$处的电压)和一个指数项的乘积。这个指数项的指数因子通常写成$-\gamma z$的形式，所以最终有

$$\gamma = \sqrt{ZY} = \sqrt{(R + \mathrm{j}\omega L)(G + \mathrm{j}\omega C)} \tag{25}$$

为了更好地理解传播常数的意义，首先注意到一般来讲γ是个复数。因此我们可以把γ显式表示成实数和虚数部分的和

$$\gamma = \sqrt{(R+\mathrm{j}\omega L)(G+\mathrm{j}\omega C)} = \alpha + \mathrm{j}\beta \tag{26}$$

于是

$$V(z) = V_0 \mathrm{e}^{-\gamma z} = V_0 \mathrm{e}^{-(\alpha+\mathrm{j}\beta)z} = V_0 \mathrm{e}^{-\alpha z}\mathrm{e}^{-\mathrm{j}\beta z} \tag{27}$$

当距离增加时，第一个指数项变小，它代表这条传输线的净衰减。第二个指数项的幅度大小为 1，它只影响相位。

2.6.4　γ 与传输线参数的关系

为了求出常数 α 和 β 与传输线参数的明确关系，我们要应用几个恒等式。首先回顾一下我们可以把一个复数同时表示成指数（极坐标）形式和直角坐标形式

$$M\mathrm{e}^{\mathrm{j}\phi} = M\cos\phi + \mathrm{j}M\sin\phi \tag{28}$$

这里 M 是复数的大小（模）而 ϕ 是它的相位。极坐标形式使我们能容易地计算一个复数的平方根（由欧拉公式知）

$$\sqrt{M\mathrm{e}^{\mathrm{j}\phi}} = \sqrt{M}\mathrm{e}^{\mathrm{j}\phi/2} = \sqrt{M}\cos(\phi/2) + \mathrm{j}\sqrt{M}\sin(\phi/2) \tag{29}$$

最后需要我们从大学数学中回想的是一对半角公式

$$\cos(\phi/2) = \sqrt{\frac{1}{2}(1+\cos\phi)} \tag{30}$$

及

$$\sin(\phi/2) = \sqrt{\frac{1}{2}(1-\cos\phi)} \tag{31}$$

现在 γ 是一个复数的平方根：

$$\begin{aligned}\gamma &= \sqrt{ZY} = \sqrt{(R+\mathrm{j}\omega L)(G+\mathrm{j}\omega C)} \\ &= \sqrt{(RG-\omega^2 LC) + \mathrm{j}\omega(LG+RC)}\end{aligned} \tag{32}$$

利用我们的恒等式并做几次变换，得到

$$\alpha = \sqrt{\frac{1}{2}\left[\sqrt{\omega^4(LC)^2 + \omega^2[(LG)^2+(RC)^2] + (RG)^2} + (RG-\omega^2 LC)\right]} \tag{33}$$

及

$$\beta = \sqrt{\frac{1}{2}\left[\sqrt{\omega^4(LC)^2 + \omega^2[(LG)^2+(RC)^2] + (RG)^2} - (RG-\omega^2 LC)\right]} \tag{34}$$

最后两个表达式显得很臃肿，但它们就是这样。如果 RG 的积比其他项小，那么我们可以作相当的简化。在这一情形下，衰减常数可以写成

$$\alpha \approx \sqrt{\frac{1}{2}\left[\sqrt{\omega^4(LC)^2 + \omega^2[(LG)^2+(RC)^2]} - \omega^2 LC\right]} \tag{35}$$

上式经过一些忽略，可进一步简化成

$$\alpha \approx \frac{R}{2}\sqrt{\frac{C}{L}} + \frac{G}{2}\sqrt{\frac{L}{C}} \tag{36}$$

这可以进一步近似

$$\alpha \approx \frac{R}{2}\sqrt{\frac{C}{L}} + \frac{G}{2}\sqrt{\frac{L}{C}} \approx \frac{R}{2Z_0} + \frac{GZ_0}{2} \tag{37}$$

于是只要单位长度的电阻与 Z_0 相比很小，且单位长度的电导与 Y_0 相比很小时，单位长度的衰减将会很小。

现在把我们的注意力转到 β 的公式中，我们有

$$\beta = \mathrm{Im}[\gamma] \approx \omega\sqrt{LC} \tag{38}$$

在损耗为零（G 和 R 均为零）的极限情况下，这些表达式简化为

$$\alpha = \mathrm{Re}[\gamma] = 0 \tag{39}$$

$$\beta = \mathrm{Im}[\gamma] = \omega\sqrt{LC} \tag{40}$$

所以一根无损线并不衰减（这毫不奇怪）。由于衰减在所有频率下是相同的（即为零），所以一条无损线**没有任何带宽的限制**。此外，传播常数的虚数部分完全与频率成正比。因为一个系统的延时等于相位对频率的（加一个负号）导数，所以一条无损线的延迟是一个常数而与频率无关：

$$T_{\mathrm{delay}} = -\frac{\partial}{\partial\omega}\phi(\omega) = -\frac{\partial}{\partial\omega}(-\beta z) = \sqrt{LC}\,z \tag{41}$$

我们现在理解了在引论中曾提到的分布系统的极佳的特性：即电容和电感并不会直接引起带宽的减少，它们只影响传播延时。设想如果我们增加单位长度的电感或电容，那么延时就会增加而带宽（理想地）却不会改变。这一特性与我们对低阶集总元件网络观察到的特性是完全不同的。

同时与低阶集总网络完全不同，传输线可以像这里看到的那样显示出与频率无关的延时。这一特性是极为希望的，因为它意味着一个信号的所有傅里叶分量将被延迟完全相同的时间，因而脉冲的形状将会保持。我们刚才已经看到了一条无损线具有这一零色散的特性。然而因为所有实际的传输线都会有非零的损耗，我们是否在实际中必须接受这一色散（或非均匀延时）呢？幸运的是正如 Heaviside⑧ 第一个指出的那样，回答是“不”。如果对传输线的常数进行某些控制，那么即使是一条有损线，仍然可以得到均匀的群时延（至少在原理上如此）。特别是，Heaviside 发现选择 RC 等于 GL（或等同地，选择串联阻抗 Z 的时间常数 L/R 等于并联导纳 Y 的时间常数 C/G）可以得到常数的群时延。此时的衰减自然是不为零（真遗憾，不能消除这一点），但常数的群时延意味着脉冲在沿线向前传播时只是幅度变小，但并不变形（即并不发生色散）。

要证明 Heaviside 是正确的并不太难。在 α 和 β 的确切表达式中令 RC 和 GL 相等，得到

$$\alpha = \mathrm{Re}[\gamma] = \sqrt{RG} \tag{42}$$

和

$$\beta = \mathrm{Im}[\gamma] = \omega\sqrt{LC} \tag{43}$$

注意 β 的表达式与无损线的一样，因此同样得到与频率无关的延时。

尽管衰减不再是零，但它仍然与频率无关；只要我们选择 $L/R = C/G$，带宽仍然是无穷大。而且特征阻抗在所有的频率下精确地等于 $\sqrt{L/C}$，而不只是在高频时才逐渐逼近这个值。

---

⑧ 此外，他还是第一个采用向量微积分把麦克斯韦尔方程写成现代形式，他也是将拉普拉斯变换引入用来求解电路问题的人。

使 $LG=RC$ 最好通过增加 L 或 C，而不是增加 R 或 G 来实现，这是因为后一种策略会增加衰减(我们假设这是不希望的结果)。哥伦比亚大学的 Michael Pupin 顺着 Heaviside 研究的思路，建议沿电话传输线周期地添加集总电感以减少信号色散。这样一种“Pupin 线圈”在 1920 年及 1930 年代显著改善了电话通信[⑨]。

2.7　有限长度传输线的行为特征

现在我们已推导了无限长传输线的许多重要特性，是时候来考虑当有限长的传输线被随意阻抗所终断时将会怎样。

2.7.1　终端匹配的传输线

一条无限长传输线策动点处的阻抗就是 Z_0。假设我们把这条线在某处切断，丢弃掉无穷长的部分，并且用单个的集总阻抗值 Z_0 来代替它，那么策动点的阻抗也必定仍然是 Z_0；对测量设备而言没有任何办法可以区分它是集中参数阻抗还是它所替代的那部分线的阻抗。因此施加到这条线的信号只是沿这条线的有限部分向下传播，最终到达这个电阻，使它变热，对全球变暖贡献了一份子。

2.7.2　终端接上任意负载阻抗的传输线

一般的，一根传输线的终端处的阻抗不会精确地等于它的特征阻抗。沿线向下传播的信号(自然地)一直保持着电压与电流的比率等于 Z_0，直到它遇到负载阻抗。然而终端负载阻抗强迫电压与电流的比率变成它自己特定的比。因此唯一能使这一矛盾化解的是使信号的一部分反射回信号源。

为了区分向前(入射)量和反射量，我们将分别运用下标 i 和 r。如果 E_i 和 I_i 是入射电压和电流，那么很显然，

$$Z_0 = \frac{E_i}{I_i} \tag{44}$$

在这根线的负载端，阻抗的不匹配引起了反射电压和电流。由于我们仍然有一个线性系统，所以系统中任何点处的总电压是入射电压和反射电压的叠加。类似地，净电流也是入射电流和反射电流的叠加。但是因为电流的这两个分量是沿相反的方向传播的，所以这里的叠加实际上就是相减。于是有

$$Z_L = \frac{E_i + E_r}{I_i - I_r} \tag{45}$$

可以重写最后一个方程以明确显示它与 Z_0 的比例关系

$$Z_L = \frac{E_i + E_r}{I_i - I_r} = \frac{E_i}{I_i}\left[\frac{1 + E_r/E_i}{1 - I_r/I_i}\right] = Z_0\left[\frac{1 + E_r/E_i}{1 - I_r/I_i}\right] \tag{46}$$

在传输线负载端反射量与入射量的比称为 Γ_L，它一般为一个复数。运用 Γ_L，Z_L 的表达

---

⑨ 然而，采用集总电感会引入对带宽的限制，而这种限制对于真实的分布传输线是没有的。因为带宽和信道容量是密切相关的，所以所有的 Pupin 线圈(它曾以花巨资装上去)最终都不得不都被移去(也是耗资巨大)以增加每条线所能运载的通话数。我们会在研究滤波器设计时，进一步探讨这个主意。

式变为

$$Z_L = Z_0\left[\frac{1+E_r/E_i}{1-I_r/I_i}\right] = Z_0\left[\frac{1+\Gamma_L}{1-\Gamma_L}\right] \tag{47}$$

求解 Γ_L 得到

$$\Gamma_L = \frac{Z_L - Z_0}{Z_L + Z_0} \tag{48}$$

如果负载阻抗等于传输线的特征阻抗，那么反射系数为零。如果一条线的终端是短路或开路，那么反射系数的幅值将为 1，这是反射系数的幅值能达到的最大值（这是对纯无源系统而言的，例如我们现在所讨论的这一个系统）。

我们可以把反射系数的概念一般化，即其是沿传输线在任意点处反射量与入射量的比：

$$\Gamma(z) = \frac{E_r e^{\gamma z}}{E_i e^{-\gamma z}} = \frac{E_r}{E_i}e^{2\gamma z} = \Gamma_L e^{2\gamma z} \tag{49}$$

按惯例，式中将传输线的负载端处的位置定义为 $z=0$，而把驱动源设在 $z=-l$ 的位置。根据这一惯例，沿传输线在任何点 z 处的电压和电流可以表示为

$$V(z) = V_i e^{-\gamma z} + V_r e^{\gamma z} \tag{50}$$

$$I(z) = I_i e^{-\gamma z} - I_r e^{\gamma z} \tag{51}$$

与通常一样，在任何点 z 处的阻抗简单地只是电压与电流的比：

$$Z(z) = \frac{V_i e^{-\gamma z} + V_r e^{\gamma z}}{I_i e^{-\gamma z} - I_r e^{\gamma z}} = Z_0\left[\frac{1+\Gamma_L e^{2\gamma z}}{1-\Gamma_L e^{2\gamma z}}\right] \tag{52}$$

替换掉 Γ_L 并且经过一系列的变换得到

$$\frac{Z(z)}{Z_0} = \frac{\frac{Z_L}{Z_0}(e^{-\gamma z}+e^{\gamma z}) + (e^{-\gamma z}-e^{\gamma z})}{\frac{Z_L}{Z_0}(e^{-\gamma z}-e^{\gamma z}) + (e^{-\gamma z}+e^{\gamma z})} \tag{53}$$

把上式写成更为紧凑的形式，得到

$$\frac{Z(z)}{Z_0} = \frac{\frac{Z_L}{Z_0} - \tanh\gamma z}{1 - \frac{Z_L}{Z_0}\tanh\gamma z} \tag{54}$$

在衰减可以忽略的特殊情形下（正如通常假设的那样为了易于分析），得到了一个相当简化的结果

$$\frac{Z(z)}{Z_0} = \frac{\frac{Z_L}{Z_0} - j\tan\beta z}{1 - j\frac{Z_L}{Z_0}\tan\beta z} = \frac{Z_L\cos\beta z - jZ_0\sin\beta z}{Z_0\cos\beta z - jZ_L\sin\beta z} \tag{55}$$

式中 z 是实际的坐标值，它总是零或负值。

作为最后一点说明，注意到这一表达式是周期性的。自然这样的特性只是在无损线中才能严格地观察到，但只要损耗可以忽略，那么实际的传输线也有类似的特性。周期性意味着我们只需考虑传输线某个有限部分（具体说就是半波长）的阻抗特性。这一特性可以在建立史密斯圆图时加以利用，后者将在第 3 章中简略地加以讨论。

2.8　传输线方程小结

我们已经看到特征阻抗和传播常数是每单位长度串联阻抗和并联导纳的简单函数

$$Z_0 = \sqrt{\frac{Z}{Y}} = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \tag{56}$$

$$\gamma = \sqrt{ZY} = \sqrt{(R + j\omega L)(G + j\omega C)} \tag{57}$$

利用这些参数——连同反射系数的定义——使我们可以推导出一个终端为任意阻抗的有损耗传输线的策动点阻抗公式

$$\frac{Z(z)}{Z_0} = \frac{\dfrac{Z_L}{Z_0} - \tanh\gamma z}{1 - \dfrac{Z_L}{Z_0}\tanh\gamma z} \tag{58}$$

在无损(或可忽略损耗)传输线的情况下,这一阻抗表达式具有相当简单和周期性的形式,这就为讨论史密斯圆图建立了基础

$$\frac{Z(z)}{Z_0} = \frac{\dfrac{Z_L}{Z_0} - j\tan\beta z}{1 - j\dfrac{Z_L}{Z_0}\tan\beta z} = \frac{Z_L\cos\beta z - jZ_0\sin\beta z}{Z_0\cos\beta z - jZ_L\sin\beta z} \tag{59}$$

2.9　人造传输线

我们刚才已经看到由无穷小电感和电容构成的一个无限长梯形网络在无穷大带宽上具有纯实数的输入阻抗。尽管无限长的结构实现起来总是有些不方便,但总可以在有限长传输线的终端接上一个等于它的特征阻抗的负载。相对来说能量是比较容易被愚弄,因为它不能区分到底是真正的传输线还是一个等于特征阻抗的电阻,所以被恰当地终断的有限长传输线的策动点阻抗仍然与无限长传输线的相同,并且仍然也在无限的带宽上保持不变。

在有些情形中我们也许希望用一个有限集总网络来近似一个连续的传输线。这样做的目的包括:实现起来比较方便,或者是对传输线的常数有较好的控制。然而,采用有限集总近似肯定不能使这样一条人工传输线的特性在无限大的带宽上与一条理想传输线相比[⑩]。因此采用集总传输线的电路设计必须考虑这一带宽的限制。

人工传输线的一个重要应用是对延迟线进行综合生成,见图 2.4。图中采用 LC 的 L 形段来综合产生延迟线。正如在连续的情形中,策动点的阻抗就是

$$Z_{in} = \sqrt{L/C} \tag{60}$$

而每一段的延时为

图 2.4　集总参数延迟线

⑩　一个比较容易理解这点的办法是注意到当一条真正的传输线用作一个延时元件时,随着频率接近无穷大能提供无限制的相移。而一条集总传输线只可以提供有限的相移(因为它只有有限个数目的储能元件),因而只有有限个数目的极点。

$$T_D = \sqrt{LC} \tag{61}$$

集总延迟线的价值在于可以得到大的延时而不必采用例如 1km 长的同轴电缆。

2.9.1 集总参数传输线的截止频率

不同于分布传输线，集总参数传输线只在一段有限的带宽内显示出常数的实数阻抗。当频率提高时输入阻抗最终会变成纯电抗[11]，这意味着实际的功率既不会送入这条线也不会送到与该线另一端相连的任何负载中。这种行为是我们在另一章中将要考虑的一类滤波器的基础。实际上，滤波器理论的出现可以直接溯源至这种结构的有限的频率响应。

截止频率是指从该频率开始，在更高的频率上没有功率传输的频率，它可以很容易地通过一条无限的（但是采用集总元件的）LC 线的输入阻抗公式求得，这里为方便起见再次列出 2.5 节中的这一公式

$$Z_{in} = \frac{Z}{2}\left[1 \pm \sqrt{1 + \frac{4}{ZY}}\right] \tag{62}$$

式中令 $Y = j\omega C$ 及 $Z = j\omega L$，于是输入阻抗为

$$Z_{in} = \frac{j\omega L}{2}\left[1 \pm \sqrt{1 - \frac{4}{\omega^2 LC}}\right] \tag{63}$$

在足够低的频率时，根号内的项具有纯负值。开方所得到的虚数项乘以 $j\omega L/2$ 的因子时，就得到了输出阻抗的实数部分。

然而随频率提高，在根号内的项最终变为零。在这个频率或以上，输入阻抗是纯虚数，所以没有任何功率可以传递到这条线上。因此截止频率为

$$\omega_h = \frac{2}{\sqrt{LC}} \tag{64}$$

由于集总传输线的特性在不到截止频率且离之较远时就已经开始变差，所以通常必须选择截止频率比感兴趣的最高频率高出相当的数目。如果需要有好的脉冲保真度，那么满足这一要求特别重要。

在设计人造传输线时，选择 L/C 的比率以提供所要求的传输线阻抗，同时要选择足够小的 LC 乘积以提供足够高的截止频率使传输线在所希望的带宽内有近似理想的特性。如果要求一个规定的总延时，那么这前两个要求就决定了必须采用的 L 形段的最少数目的段数。

2.9.2 集总参数传输线的终断

有关如何来终断图 2.4 电路的问题始终是存在的。一种选择是如图 2.5 终断，负载是简单地接到这个电容器两端。另一种选择（见图 2.6）是以电感作为终结。尽管这两种选择多少都能工作，另一个较好的选择则是采用折中的办法，即在传输线的两端均采用接一个半段，如图 2.7 所示。

---

[11] 从对该网络的观察中应当很清楚，策动点阻抗最终变为输入电感的阻抗，因为电容器最终是起了短路作用。

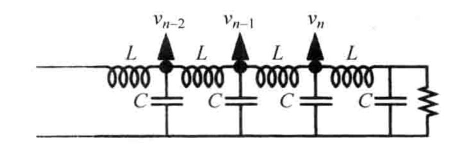


图 2.5　集总参数传输线终断的一种选择

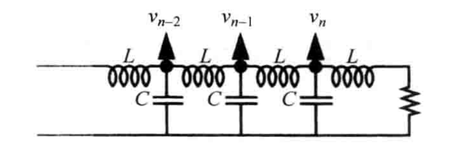


图 2.6　集总参数传输线终断的另一种选择

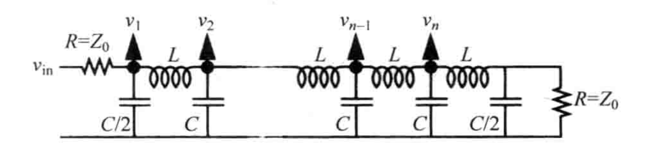


图 2.7　集总参数传输线的半节终断法

这样一种折中选择可以使带宽超过图 2.5 和图 2.6 所给出的。每半段贡献完整段时延的一半，所以在每一端各加半段就增加了一个整段的时延。而且更为重要的是一个半段的截止频率是一个整段的两倍，这就是为何能得到较好带宽的真正原因。

许多应用要求延迟元件提供相当于一个四分之一波长的传输线的近似延迟，就像在讨论诸如耦合器与合成器这类的电路单元时将会看到的。π 形段或 T 形段的拓扑都是合适的，而最常用的是三个元件组成的，见图 2.8。这两个网络具有同样的传递函数，因此可以根据实际的考虑来进行选择，比如说如何更好地兼容寄生元件。

图 2.8　用于近似 λ/4 延迟线的简单 π 形与 T 形单元网络

照例，L/C 的比率被选择为要模拟的传输线的特征阻抗的函数

$$\sqrt{L/C} = Z_0 \tag{65}$$

更进一步，不难证明当图示的这些网络为内阻为 Z_0 的信号源所驱动，并且终端负载是 Z_0，它们在下面的频率下，输出滞后输入一个 90°的相位(即正交)：

$$\omega_0 = 1/\sqrt{LC} \tag{66}$$

求解这两个元件的值，得到

$$C = 1/(\omega_0 Z_0) \tag{67}$$

$$L = Z_0/\omega_0 \tag{68}$$

2.9.3　*m* 参数推导出的半段网络

LC 半段端口的阻抗在截止频率的 30%～40%时开始明显地增加，这是由于由输出电容以及它所看到的其余电抗所构成的并联谐振所致。这一情况可以采用只是比单独一对 LC 稍微复杂一点的半段来调整。具体地说，如果电容用串联的 LC 分支来代替，那么阻抗大致保持不变的频率范围可以进一步增加，因为串联谐振分支的阻抗的减少可以帮助抵消阻抗的增加。

能实现这一所希望结果的一个简单网络如图 2.9 所示。在有关滤波器的那一章会给出更详细的推导。眼下，我们先接受下面这些公式给出的元件值

$$L_1 = \frac{mL}{2} \tag{69}$$

$$L_2 = \frac{1-m^2}{2m}L \tag{70}$$

$$C_1 = \frac{mC}{2} \tag{71}$$

图 2.9　用 m 参数推导的用于传输线终断的半节网络

用这一方法改进的网络称为 m 参数推导的半节网络，这是因为对于参数 m 的任何值，标称特征阻抗保持与简单的 LC 半段网络相同。这可以通过直接代入公式(62)来验证。

对于 m 值大约为 0.6 时，直到截止频率的约 85%的范围内阻抗都保持基本不变。因此普遍采用这一选择。这个 m 推导的半节网络可以被用作一根人工传输线的终结端的终止电路节，或者用来构造整个传输线。我们会在第 22 章讨论滤波器设计这个重要课题时重新回到这些概念的讨论中来，不过是在滤波器这个大背景下。

2.10　总结

我们已经确立了在集总和分布领域间的模糊界线并且发现集总的概念可以延伸到分布领域。在进行这样的延伸时，我们已经发现存在几个，也许还有更多的方法可以用时延而不是带宽来换取增益。作为对这一课题的最后一个观察，重申一下下面的结论或许是值得的：避免使用增益-带宽间的折中互换要求我们放弃单极点电路的动态特性，因此，我们已看到的所有用延时换取增益的结构都包含有许多个能量存储元件。观察这一问题的另一种方式是认识到如果打算用延时来换取任何东西，那么必须有能力提供大的延时。但是大的延时意味着每单位频率有较大数量的相位变化，并且如果想在一个大的带宽上工作，那么所要求的总的相位变化就会非常大。同样，对大的相移的需求意味着必然要求有许多极点(因而需要大量的电感和电容)，这就导致了我们已经见过的相对复杂的网络[12]。然而，如果追求最大可能的工作带宽，那么采用这些分布的概念就完全是必须的。在第 12 章中将会看到分布的概念也可以应用到有源电路中，以实现具有特别大带宽的放大器，这是用牺牲时延来换取改善了的带宽而实现的。

---

⑫　超再生放大器是一个例外。在这种放大器中，系统的时变特性引起了这个单级放大器响应的信号混叠(aliasing)，这个组合的效果产生了一个类似于多级这样的单级放大器级连的响应。

第3章 史密斯圆图与散射(S)参数

3.1 引言

在微波电路与系统的设计所起始的那个时代，器件和互连线尺寸通常都很大而不适合集总参数的描述。而且由于缺乏适当详细的模型以及相应的计算工具，迫使工程师们用图解的方法，把系统当作双端口的“黑匣子”来处理。在这些作图辅助手段中，最有用的当数史密斯圆图，其起源于20世纪30年代，一个以计算尺为主要计算工具的时代。尽管史密斯圆图作为一种辅助计算工具，今天比起过去来，它们已经不再那么有关了，但作为一个例子，RF仪器仍然在以史密斯圆图的形式给出数据。用史密斯圆图来目视一些运算能揭示丰富的设计直觉，现在依然如此。而现代的计算辅助手段往往不幸地忽略了这些设计直觉。本章因此提供了一个有关史密斯圆图的简略历史与推导，并解释为什么一组特别的变量(S参数)在描述双端口微波电路时要优于其他参数(例如阻抗或导纳)。

3.2 史密斯圆图

介绍史密斯圆图的有关课件往往忽略了其历史背景，这样就给学生留下了一个印象，即史密斯圆图是一下子完整地形成的。这个印象又反过来使得许多学生精神上缺乏这样的信心，即他们应该能立即领会这个圆图必然“明显”具备的不可言喻的美丽、逻辑性与强大的功能。然而，真正的故事却是史密斯圆图跨越了大约十年的一点点进步积累的结果。史密斯圆图对于史密斯本人而言也不是那么显而易见，因此或许它也不应该对我们会立即如此明显。此外，谁是那个史密斯?

Phillip Hagar Smith(史密斯)在20世纪20年代后期加入贝尔实验室，正好是在那一次股票市场大跌之前。他的职业生涯的早期是进行天线系统的工作，因此阻抗匹配的问题是经常遇到的。他发现自己不得不一遍又一遍地进行类似的计算。那种计算十分繁杂，导致他发明了一连串的图形辅助手段(记住，当时没有电子制表软件，只有计算尺)。到了20世纪30年代中期，他已经设计出阻抗的笛卡儿与极坐标表示，可以用来帮助计算。到了大约1937年，他已从贝尔实验室的同事处学到了足够多的数学(特别是保角变换的技巧)来完成他的最后几步，并且实现了基本上是现在看到的那种版本的圆图[①]。

读者现在可以感到一点自慰了，因为在那时候除了史密斯本人与他在贝尔实验室的朋

---

① R. Rhea, “Phillip H. Smith: A Brief Biography”(史密斯：一个简短的自传)(这是一个对史密斯下面这本书再版的简介：*Applications of the Smith Chart: In Waveguide, Circuit, and Component Analysis*(史密斯圆图在波导、电路与元件分析上的应用)，Noble Publishing，2000)

友，很少有人对他的圆图给予关注[②]。最终，在1939年1月《电子学》(Electronics)期刊发表了史密斯关于他的新的圆图的文章[③]。这篇文章的发表，引起了好多人的惊叹与不解，或者两者兼有。这种反应并不奇怪，在那个计算尺的时代，工程师们经常不是发明图，就是列线图解计算表(nomograph)，或者是这样那样的图解计算辅助工具。史密斯圆图必须与所有这些图形计算工具竞争，才能出头。史密斯圆图看起来吓人的外表也妨碍了人们对它普遍的接受。

这种情况因第二次世界大战前夕MIT的辐射实验室成立而得以改变。在辐射实验室(Rad Lab)开发雷达的许多物理学家并没有被作为史密斯圆图基础的数学所吓退(甚至或许是被这些数学而吸引)，因为他们在其他的场合下已经遇到过。他们对数学的习惯，加上他们手头任务的紧迫性，使得史密斯圆图很快就被接受为一个不可或缺的设计工具。在二次大战结束前，众多的工程师已经变得十分熟悉史密斯圆图了。史密斯本人在1944年发表的另一篇文章[④]又帮助了圆图的进一步推广。

史密斯圆图的数学基础是双线性变换(两个线性函数的一个比率)，它将反射系数与归一化的负载阻抗 Z_{nL} 联系起来：

$$\Gamma = \frac{\dfrac{Z_L}{Z_0} - 1}{\dfrac{Z_L}{Z_0} + 1} = \frac{Z_{nL} - 1}{Z_{nL} + 1} \tag{1}$$

这个归一化的负载阻抗和 Γ 之间的双线性关系，也即两者是一一对应的：知道一个相当于知道另一个。这个观察的结论是很重要的，因为我们所熟悉的、画在史密斯圆图上的曲线，只不过是在 Γ 平面上画出的常数电阻和电抗的等值线。

一个自然要问的问题是既然把阻抗的实数和虚数部分直接画在标准的笛卡儿坐标中更容易，为什么还要如此麻烦地把阻抗画在一个非直线正交(nonrectilinear)的坐标系统中呢？这一不是显而易见的选择至少有两个很好的理由。一个理由是若试图在一个有限大小的纸上画一个无穷大的阻抗会直接引起许多明显的实际问题。相反如果画 Γ 则能很好地处理任意大小的阻抗，因为对于无源负载 $|\Gamma|$ 不可能超过1。另一个理由是当一条无损传输线的终端是一个固定的阻抗时，Γ 则每半个波长重复一次，因此画 Γ 是表示这一周期性特性的一种自然而又简洁的编码方法。采用史密斯圆图进行计算的优越性主要来自这一简洁的编码表示方法，例如它可以使工程师们迅速确定把一个阻抗转换成一个特定值所需要的传输线长度，或者通过目测就可以读出反射系数的大小。

在公式(1)中阻抗和 Γ 之间的关系可以看成是把一个复数映射成另一个复数。在这一情形中，我们已经注意到这是一种被称为**双线性**(*bilinear*)**变换**的一种特殊映射，因为它是两个线性函数的比。在双线性变换的许多性质中，一个特别有用的是一个圆在映射后仍然是圆。在这个上下文中，一条直线可以看成是一个具有无穷大半径的圆。因此，圆和直线被映射成圆或者直线。

---

② Louis Smullin，私人通信

③ “A Transmission Line Calculator”(一个传输线的计算器)

④ “An Improved Transmission Line Calculator”(一个改进了的传输线计算器)，*Electronics*(《电子学》)，v. 17, January 1944, p. 130.

借助于公式(1)，可以很容易地显示 Z 平面的虚轴被映射成 Γ 平面上的单位圆，而在 Z 平面上电阻为常数的其他直线被映射成不同直径的圆，这些圆都在点 $\Gamma=1$ 处相切，见图 3.1。

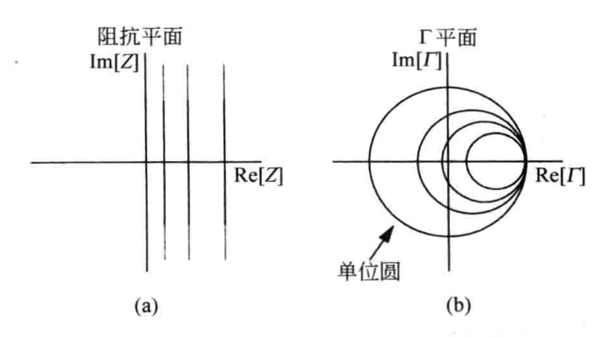


图 3.1　在 Z 平面上的等电阻值线映射到 Γ 平面上的圆

在 Z 平面上，等电抗的直线与等电阻的直线正交，而且这一正交性在映射以后仍然被保持(就像它们在所有的保角变换时那样)。由于直线映射成直线或圆，我们可以预期等电抗直线被转换成如图 3.2 所示的圆弧。史密斯圆图只不过是简单地由在 Γ 平面上等电阻和等电抗的等值线组成，但没有明确画出 Γ 平面的坐标轴而已。

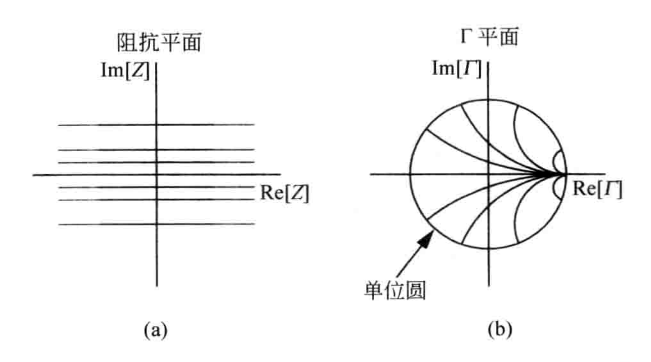


图 3.2　Z 平面上的等电抗的直线映射成 Γ 平面上相应的等值轨迹

正如前面提到的，由于现在史密斯圆图的主要用途是作为一种标准的方式来呈现阻抗(或反射系数)的数据，所以值得花一点时间来熟悉它一下。史密斯圆图的中心相应于零反射系数，因此它表示一个用于归一化阻抗的电阻值。

Z 平面的下半平面被映射成在 Γ 平面上单位圆的下半圆，因此电容性阻抗总是处在那些位置。类似地，Z 平面的上半部相应于单位圆的上半圆，那里是电感性阻抗的位置。逐渐变小的等电阻圆对应于逐渐增大的电阻值。$\Gamma=-1$ 点相应于零电阻(或电抗)，而 $\Gamma=1$ 点相应于无穷大电阻(或电抗)。

举一个简单但具体的例子，让我们画出一串联 RC 网络的阻抗的图，其电阻为 100Ω，电容为 25pF，都归一到 50Ω 的系统。因为阻抗是实数部分(等于电阻)和虚数部分(等于容抗)的和，所以在 Γ 平面上相应的轨迹必定是沿 $R=2$ 的等电阻的圆。电抗部分的值从 DC 时的负无穷大变化到在无穷大频率时的零。由于它的符号总是负的，所以轨迹必然就

是那个等电阻圆的下半圆，且随着频率的增加，它从 $\Gamma=1$ 沿顺时针方向变化，如图 3.3 所示。

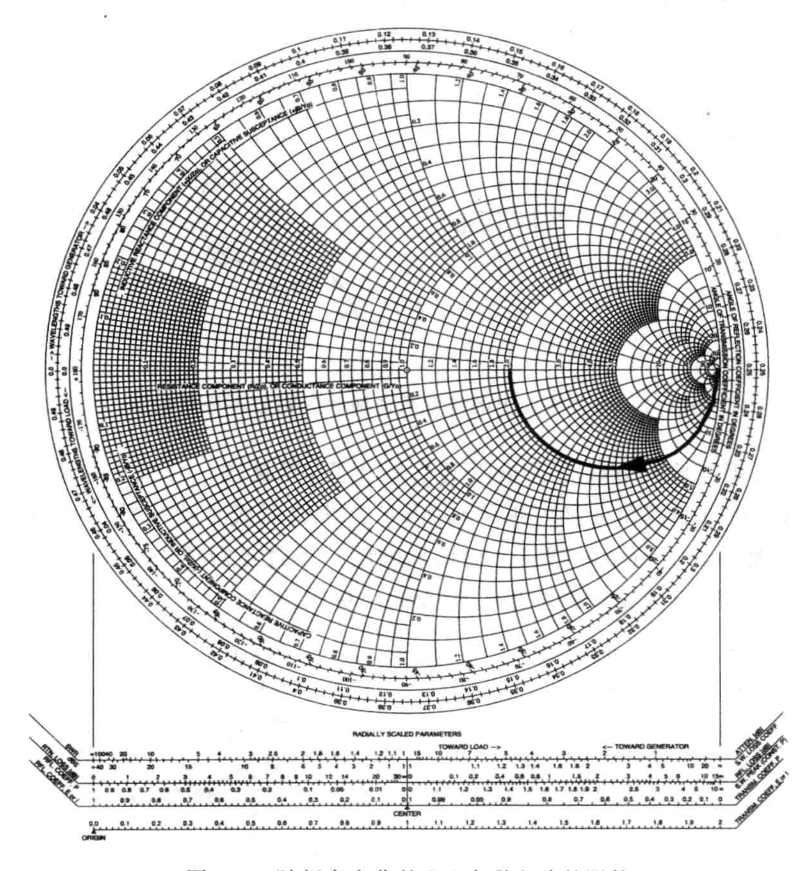


图 3.3　随频率变化的 RC 串联电路的阻抗

注意这条曲线是对任意值的一个电容与 100Ω 电阻串联的阻抗轨迹。当电容值发生变化时，只是对应于轨迹上某一特定点上的频率发生了变化。由此可以得到一个推论：对于固定的频率，增加(或减少)电容值仅仅改变了沿着这同一半圆圆弧的位置。所有这些曲线(即不同电阻值，译者注)在频率等于零时汇聚于对应于无穷大阻抗的共同点(即 $\Gamma=1$ 这个点，译者注)。

一个与同样的 100Ω 电阻串联的电感会列遍一个圆的上半圆。又一次，顺时针的遍列相应于增加的频率。就像与 RC 的例子那样，电感值的变化在给定频率的情况下只是简单地导致沿半圆周的滑动。所有的串联 RL 网络的阻抗轨迹都共享一个在无穷大频率下的无穷大阻抗的共同点。

在前面两个例子的基础上，认识到一个串联 RLC 网络在频率变化时沿一个(完整的)圆周滑动就会是什么了不起的飞跃。在串联谐振时，阻抗是一个纯实数(电阻)，这对应于水平轴上的一个点。

史密斯圆图用来求并联阻抗的值也很方便。我们可以同样容易地用导纳或阻抗来表示反射系数：

$$\frac{Z_{nL}-1}{Z_{nL}+1}=\frac{1/Y_{nL}-1}{1/Y_{nL}+1}=-\frac{Y_{nL}-1}{Y_{nL}+1} \tag{2}$$

除了有一个符号的变化，用阻抗与导纳表示的反射系数是一样的。这样，等电导的轨迹也是一些圆，而等电纳的轨迹是与等电导的圆相正交的圆弧。式(2)中的负号很容易通过将圆图如图 3.4 所示的依圆心不动对称反转过来而得到。在如图所示的朝向下，电感性的电纳是在下半圆，而电容性的电纳是在上半圆。某些史密斯圆图的商用版本上，阻抗与电纳的等值线都有画出(经常用不同的颜色或不同的灰度表示)。这个组合复杂得有点吓人，除非对史密斯圆图熟悉到一定程度。

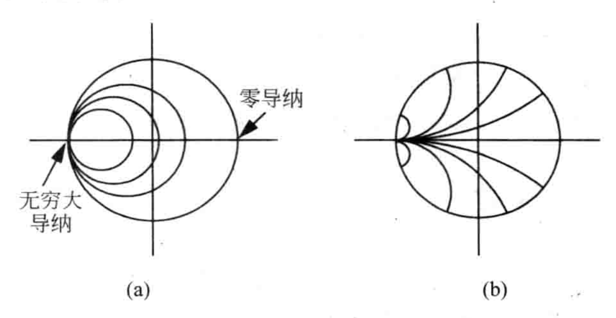


图 3.4　在导纳圆图上的等电导(a)与等电纳轨迹

一个并联的 GC 网络当频率从零变到无穷大时，周历电纳圆图的整个上半圆。所有这种曲线在无穷大电纳的点处相切，而且在无穷大频率下汇聚在这个点。类似地，一个并联的 GL 网络周历在圆图下半部的半圆周。又一次，所有这样的半圆在导纳等于无穷大处相切，而且在频率等于零时汇聚在该点。一个并联的 GLC 网络则在频率变化时，周历一个完整的圆周。随着频率的增加，点的移动是逆时针的，这与串联阻抗图的顺时针移动正好相反。

除了帮助图解增加并联与串联的元件是如何改变阻抗与导纳之外，史密斯圆图也使我们能快速地计算沿传输线的阻抗变化。仔细看一下商用的史密斯圆图，可以发现这样的短句，如“朝向信号源的波长”与“朝向负载的波长”。这些短句分别被配以逆时针与顺时针指向的箭头。其解释是十分简单的：想象一下一根无损耗的传输线被某个阻抗所终断。假若我们朝着信号源移动某个量，我们知道反射系数会以某种方式发生变化。具体地说，

$$\Gamma(z)=\Gamma_{L}e^{2\gamma z} \tag{3}$$

这样可以得出结论，反射系数的大小并不会发生改变，仅仅是相位角变了。现在史密斯圆图是将阻抗变换成反射系数(Γ)的结果。这样，Γ 实际上是画在笛卡儿坐标上，但是相应的坐标轴并没有印在普通的史密斯圆图上(然而，这里为了说明方便，我们包括了这些轴)。但是，假装这些坐标轴是在圆图上，你可以看到 Γ 是如何能够被很快地读出来。史密斯圆图的中心对应于完美的匹配，而且是 Γ 的不可见的直角坐标系统的原点。常数 Γ 的等值线因此就是以圆心为坐标原点的圆，失配越大(即 Γ 幅度越大)，其半径也越大。因此，当我们沿传输线移动时，我们只是跨过与单位圆同心的圆的一部分圆弧。这就使得我们很容易计算在传输线上任一点的阻抗。参见图 3.5。

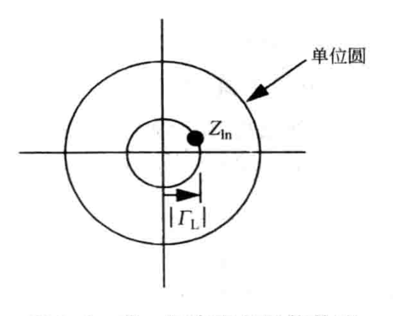


图 3.5　单一频率下的阻抗轨迹

史密斯圆图不仅在解析分析上有用，它在推断出为产生特定的阻抗变换所需的传输线的长度上也有极大的价值。这一点会在第 4 章中进一步说明。一个对应于最大的可容许的 Γ 的具有常数反射系数幅度的圆可以被叠加在史密斯圆图上，以便于确定匹配的目标是否已达到。

作为进一步的说明，一个典型的史密斯圆图就在圆图的下方专门有一组额外的度量尺。这些尺子是经反射系数校准过的，或是经将反射系数变换到阻抗失配的其他度量：驻波比(SWR)与回波损失(RL)，校准过的。

史密斯圆图还有非常多的其他性质，用这些性质来图解与快速地进行计算的类型真的是很多。我们将在下一章中研究一个性质(但这是一个十分重要的性质)：阻抗匹配。因为计算机的计算已经大大减小了史密斯圆图的作用，我们仅给出一个不完整的描述，有兴趣的读者可以参考史密斯 1944 年的文章来找到更进一步的应用[⑤]。

3.3 散射参数(S 参数)

工程师们已经提出了许多(也可能是太多了)描述系统的方法。为了简化分析或者是为了阐明重要的设计准则，宏观描述通常很有价值，它保留了系统的输入-输出特性但不涉及系统内部结构的细节。在较低频率时，最常用的表示方法是采用阻抗或导纳参数，或者这两者的混合(更明确地称为混合参数)。

线性两端口表征方法的基础就是在一个端口的电压(或电流)可以被表示为在这个端口上的电流(或电压)与在另一端口上的电压或电流的线性组合。线性组合中的系数可以是没有量纲的，也可以有阻抗或导纳的量纲，这取决于所选择的变量。比如说，在阻抗表征中，我们用端口电流表示端口电压。对于图 3.6 所示的双端口网络，相关的方程为

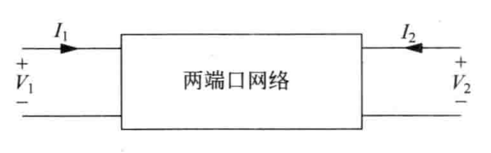


图 3.6 端口变量的定义

$$V_1 = Z_{11}I_1 + Z_{12}I_2 \tag{4}$$

$$V_2 = Z_{21}I_1 + Z_{22}I_2 \tag{5}$$

最方便的是使端口逐次开路(即电流等于零，译者注)通过实验来确定各种 Z 参数，因为此时许多项变成了零。例如当输出端口开路时确定 Z_{11} 是最简单的，因为在这一条件下方程(4)中的第 2 项为零。此时用一个电流源驱动输入端口并测量在输入端口所产生的电压就可以直接计算出 Z_{11}。类似地，使输入端口开路，用一个电流源驱动输出端口，此时测量 V_1 就可以确定 Z_{12}。

短路条件用来确定导纳参数，而联合开路和短路条件可以确定混合参数。在低频条件下这些表征系统的方法被普遍采用，这可直接归因为很容易用实验来确定这些参数。

⑤ 关于史密斯圆图其他人已经写了许多书与文章，但你也可以直接从史密斯本人学到。他的 1944 年的文章描述的圆图性质，其实就是我们现在所知道的。对这个课题的更透彻的处理，见他的书 *Electronic Applications of the Smith Chart: In Waveguide, Circuit, and Component Analysis*. McGraw-Hill, New York, 1969(2000 年 Noble Publishing 重印)。

然而在高频时，提供合适的短路或开路是相当困难的，特别是要在一个很宽的频率范围内实现这些条件。而且，有源高频电路工作时的阻抗常常是相当模糊的，它们在端口开路或短路时可能发生振荡或者甚至停止工作，因此需要一组不同的两端网络参数来避开这些实验问题。这组参数称为散射参数（或简称为 S 参数），它们利用了这样一个事实，即当一条传输线为它的特征阻抗终止时不会引起任何反射[⑥]。因此在仪器和被测系统之间的互连线可以采用任意我们认为方便的长度，因为不必提供短路或开路，这就大大简化了测试仪器的设置。

正如上面所指的，端口开路或短路对于低频双端口网络的描述是很方便的，因为许多项变为零，从而简化了数学运算。散射参数保留了这一所希望的性质，它用入射和反射（散射）电压波而不是用端口电压或电流（在高频时它们反正是很难唯一地被定义的）来定义输入和输出变量。

如在图 3.7 中可以看到的，源端口和负载端口都为 Z_0 所终断。当输入和输出变量如图所示那样定义时，两个端口的关系可以写成：

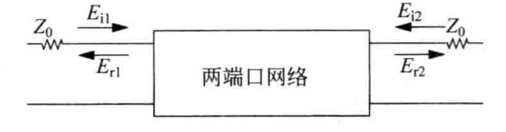


图 3.7　S 参数端口变量的定义

$$b_1 = s_{11}a_1 + s_{12}a_2 \tag{6}$$

$$b_2 = s_{21}a_1 + s_{22}a_2 \tag{7}$$

式中

$$a_1 = E_{i1}/\sqrt{Z_0} \tag{8}$$

$$a_2 = E_{i2}/\sqrt{Z_0} \tag{9}$$

$$b_1 = E_{r1}/\sqrt{Z_0} \tag{10}$$

$$b_2 = E_{r2}/\sqrt{Z_0} \tag{11}$$

用 Z_0 的平方根来归一是很方便的，因为它使各个 a_n 和 b_n 值的平方等于相应的入射或反射波的功率。

当输出端口的终端负载为 Z_0 时，驱动输入端口便使 a_2 置于零，这使我们可以确定以下参数：

$$s_{11} = \frac{b_1}{a_1} = \frac{E_{r1}}{E_{i1}} = \Gamma_1 \tag{12}$$

$$s_{21} = \frac{b_2}{a_1} = \frac{E_{r2}}{E_{i1}} \tag{13}$$

因此 s_{11} 就是输入反射系数，而 s_{21} 是一种增益，因为它将输出波与输入波联系起来。它的幅度的平方值是一种功率增益，称为源端口和负载端口阻抗均为 Z_0 时变送器的正向功率增益。我们将其确切定义延迟在附录 A 里讨论（见 3.4 节）。

类似地，如果把输入端口作为终端而驱动输出端口则得到：

$$s_{22} = \frac{b_2}{a_2} = \frac{E_{r2}}{E_{i2}} = \Gamma_2 \tag{14}$$

⑥ K. Kurokawa, "Power Waves and the Scattering Matrix"（功率波与散射矩阵）*IEEE Trans. Microwave Theory and Tech.*, v. 13, March 1965, pp. 194-202.

$$s_{12}=\frac{b_1}{a_2}=\frac{E_{r1}}{E_{i2}} \tag{15}$$

式中我们注意到 s_{22} 是输出反射系数；s_{12} 为反向传输系数，它的数值的平方就是源端口和负载端口阻抗均为 Z_0 时变送器的反向功率增益。

一旦一个双端口网络的特性用 S 参数来表征，那么原则上不必知道这个双端口网络内部工作的任何情形就可以直接设计系统。例如，增益、噪声系数和稳定性判据就可以重新用 S 参数来表示⑦。然而应当记住，一种宏观方法必然会丢弃掉一些可能很重要的信息，如对参数或工艺变化的灵敏度。由于这一原因，S 参数测量常用来计算已知拓扑中的元件值，这个拓扑是根据第一原理或物理考虑而确定的。

总之，S 参数之所以成为高频研究中几乎是无处不在的，是因为连接设备的信号电缆的长度不一定非得是"零"，也不需要实现一个短路或开路。在双端口网络的两边接上 Z_0 的端口阻抗可以大大减少振荡的可能性。

3.4 附录 A 对于单位的注解

由于在分布系统中很难唯一地确定电压和电流，也由于 RF 工程师们偏爱考虑功率增益，这使得功率很自然成为一个在 RF 电路与系统中特别受关注的量。自然，功率的大小是用 W 来表示，但使行外人士混淆和出错的是各种各样的分贝表示法。例如经常用到的"dBm"，这里的"m"是指 0dB 的参照功率是 1mW，而"dBW"则表示参照功率为 1W。如果参照阻抗值为 50Ω，那么 0dBm 就相当于电压有效值(rms)约为 223mV。

尽管这些定义已非常清楚，但还是有些人热衷于坚持混用"伏特"和"瓦"，他们重新定义 0dBm 的含义为 223mV 而不管阻抗有多大。这一重新定义不仅没有必要，因为我们总可以定义一个 dBV，而且也是非常容易出错的。正如我们将要看到的，一些关键的性能指标测量——如线性度和噪声系数——是与确切的功率比紧密相关的，特别是在研究级连系统的情况下。混淆功率比与电压比将导致整个出错。因此在全书中，0dBm 总是确切地指 1mW，而 0dBV 则指 1V。

已经解决了上面这些易混淆的定义，我们再来看其他一些定义。在讨论振荡器或功率放大器中的噪声或失真乘积项时常用到"dBc"，这里"c"指 0dB 的参照量是载波功率(c 代表 carrier，译者注)。

大多数工程师熟悉工程性的前缀，从 10^{-12}(pico，皮)到 10^{12}(tera，太)。补充比皮还要小的量有 femto(f，飞，10^{-15})，atto(a，阿，10^{-18})，zepto(z，10^{-21})，与 yocto(y，10^{-24})，这里面的某些听起来有点像马克思的兄弟们(作者的开玩笑话，译者注)。在'太'之上则有 peta(P，10^{15})，exa(E，10^{18})，zetta(Z，10^{21})及 yotta(Y，10^{24})。读者可以看到与正的指数相关联的前缀的缩写都是大写的(一个例外是 kilo(千)，这是为了避免与绝对温度的单位 K 发生混淆)，而指数为负的则用小写(对 micro 是用 μ，这是避免与 milli(毫)发生冲突)。有了这些额外的前缀，我们可以表示跨度为另外 24 个数量级的量。应该说，有了 48 个数量级，对大多数应用

⑦ 有关这一内容有代表性的参考书是 G. Gonzalez 所著的 *Microwave Transistor Amplifiers*，第二版，出版社：Prentice-Hall，Englewood Cliffs，NJ，1997.

也就够了。

根据国际标准，如果一个单位是从人名得来的，那么只有简写是用大写字母(瓦特是 W 不是 Watt；电压是 V 不是 Volt，等等)[8]。举个例子来说，这种选择保证了"two Watts"是指 Watt 家的两个成员，而不会是每秒 2J。绝对温度用开尔文(kelvins)(而不是用冗余的 kelvin 度)，且简写成 K。英升(liter 或 litre)是个例外，它的简写可以为 l，也可以为 L，后一个用法主要是为了避免与数字 1 的混淆。

最后，在微波工程师之间可以引起争论的一个很好的话题是关于词头"giga"怎么发音。这个词头起源于希腊语，两个"g"均应像"giggle"一词中的"g"那么发音，这一选择为 ANSI(美国国家标准协会)和 IEEE(国际电气与电子工程师协会)都认可。但仍然有相当多的一部分人把第一个"g"发音成像"giant"一词中的"g"那样。这个关于在 giga 中"g"的发音方式各有其拥护者，他们固执自己观点的程度远远超出了问题本身的重要性。你可以在下一次微波工程师的聚会上试试问与会者的观点，其争论之激烈，可能只能靠扔史密斯圆图来解决(作者的笑话，译者注)。

功率增益的定义

既然我们在讨论定义的话题，我们应当来讨论一下"功率增益"。你可能想，这个定义很明确啊，没有什么歧义。这是可以理解的，但是你依然可能是错的。在微波领域里，存在着 4 种人们经常遇到的功率增益。搞清楚哪个是哪个，十分重要。

通常的你所熟悉的功率增益，就是以你熟悉的方式定义的。它是实际上提供给负载的功率，除以由信号源实际给出的功率。然而，信号源驱动的负载阻抗可能并不知道(特别是在高频下)，实际上测量这个量是十分困难的。因而，其他的功率增益定义就出现了。

传送功率增益(Transducer power gain)(这是本章中用的术语)是传给负载的功率除以信号源最大可提供(available)提供的功率。假若源与负载的阻抗都是标准化了的值，那么计算提供了的功率与最大可提供的功率是直截了当的，这样就避开了前面那一段落说明的测量的困难。从定义的角度来讲，如果系统的输入阻抗正好是信号源阻抗的复数共轭，那么传送功率增益与功率增益是相等的。

最大可提供的功率增益或称资用功率增益(*available power gain*)是在系统的输出端最大可能得到的功率，除以由源可能提供的最大功率。插入功率增益(*insertion power gain*)是由系统实际上传送到负载上的功率，除以那个当信号源直接与负载相接时(即，不经考虑中的系统，译者注)的功率。任何一种上面提到的功率增益定义可能会用到，这取决于具体的场合。

最后是，如果输入、输出端口都是匹配的话，那么功率增益的这 4 种定义就完全相同了。也只有在这个场合，才可以放心地讲功率增益，而不必作任何具体的解释。

3.5　附录 B　为什么采用 50Ω 或 75Ω

大多数射频仪器和同轴电缆都具有标准阻抗 50Ω 或 75Ω。由于这两个阻抗值到处都能遇到，所以容易推想这两个值大概有些"神圣"的味道，并且会想应当在所有的设计中都用

---

⑧ 分贝(decibel)是根据 Alexander Graham Bell(贝尔)命名的，因此是 dB，而不是 db。

到它们。在这一附录中我们首先解释一下这两个数字是怎么最初被采用的，以便了解什么时候要用到这两个阻抗值，什么时候并无必要。

3.5.1 功率处理能力

考虑一根以空气作为电介质的同轴电缆。必然存在着某个电压会使电介质击穿。对于内导体直径固定的情形，我们可以加大外直径来提高击穿电压。然而特征阻抗因此就会加大，这又会减少传送到负载上的功率。由于这两个相互竞争的因素，所以存在一个最优的导体直径之比，使得同轴电缆的功率传输能力最大。

在确定了存在这一最大值之后，我们需要推导出几个公式来计算达到这个最大值所要求的实际尺寸。具体来说，我们需要一个公式计算在两个导体间最大的电场强度，以及另一个公式计算一根同轴电缆的特征阻抗：

$$E_{\max} = \frac{V}{a\ln(b/a)} \tag{16}$$

$$Z_0 = \sqrt{\frac{\mu}{\varepsilon}} \cdot \frac{\ln(b/a)}{2\pi} \approx \frac{60}{\sqrt{\varepsilon_r}} \cdot \ln\frac{b}{a} \tag{17}$$

式中 a 和 b 分别是内外导体的半径；ε_r 是相对介电常数，对我们这里以电介质为空气的导线，它实际上就是 1。

下一步是认识到传送到负载的最大功率正比于 V^2/Z_0。利用我们的公式，就得到

$$P \propto \frac{V^2}{Z_0} = \frac{E_{\max}^2 \cdot a^2[\ln(b/a)]^2}{(60/\sqrt{\varepsilon_r}) \cdot \ln(b/a)} = \frac{\sqrt{\varepsilon_r}\,[E_{\max}^2 \cdot a^2\ln(b/a)]}{60} \tag{18}$$

下一步求导，并令其为 0，我们希望由此求出的是最大值而不是最小值：

$$\frac{\mathrm{d}P}{\mathrm{d}a} = \frac{\mathrm{d}}{\mathrm{d}a}\left[a^2\ln\left(\frac{b}{a}\right)\right] = 0 \Rightarrow \frac{b}{a} = \sqrt{\mathrm{e}} \tag{19}$$

将上式求出的比率代回到我们的公式中求特征阻抗，得到了它的值为 30Ω。这就是说，为了使一根外径一定的空气介质的传输线具有最大的功率传输能力，我们希望选择尺寸使 Z_0 等于 30Ω[9]。

但是等一下：甚至对 30 这个较大的值，也不能说它就是等于 50。所以显然我们还没有回答我们原先的问题。我们需要再考虑一个因素：电缆的(信号，译者注)衰减。

3.5.2 衰减

可以证明(但我们这里不证明它)由于电介质损耗引起的每单位长度的衰减实际上与导体的尺寸无关。如果我们简化公式只考虑由电阻损耗引起的衰减，得到：

$$\alpha \approx \frac{R}{2Z_0} \tag{20}$$

式中 R 是每单位长度的串联电阻。在足够高的频率时(我们将很快考虑这一情形)，R 主要是由于趋肤效应引起的。为了减小 R，我们希望加大内部导体的直径(以得到更多的“皮肤”面积)，但这同时又往往会减小 Z_0，因此我们不清楚如何才能达到目的。我们再一次看到了

[9] 史密斯因为这个结果还被授予一个专利。这个专利的唯一认可的贡献是明确地提出了这个比率，$\sqrt{\mathrm{e}}$(1.6487)。

在两种相反效应之间的竞争，并且我们可以再次指望最优值会发生在 b/a 的某一特定值，也就是在某一特定的 Z_0 处。

正如上面提及的，我们将通过几个公式来求得实际的数值结果。我们这里需要的唯一的一个新公式是电阻 R 的表达式。如果我们作出一个通常的假设，即电流均匀地流过一个厚度等于趋肤深度 δ 的薄圆筒，那么可以写出：

$$R \approx \frac{1}{2\pi\delta\sigma}\left[\frac{1}{a}+\frac{1}{b}\right] \tag{21}$$

式中 σ 是导线的电导率，而 δ 总是等于：

$$\delta = \sqrt{\frac{2}{\omega\mu\sigma}} \tag{22}$$

利用这些公式，衰减常数可以表示为

$$\alpha = \frac{R}{2Z_0} \approx \frac{\frac{1}{2\pi\delta\sigma}\left[\frac{1}{a}+\frac{1}{b}\right]\sqrt{\varepsilon_r}}{2\left[60\ln\left(\frac{b}{a}\right)\right]} \tag{23}$$

求导并令其为0，但现在我们希望求出的是最小值而不是最大值，我们得到：

$$\frac{d\alpha}{da} = 0 \Rightarrow \frac{d}{da}\frac{\left[\frac{1}{a}+\frac{1}{b}\right]}{\ln\left(\frac{b}{a}\right)} = 0 \Rightarrow \ln\left(\frac{b}{a}\right) = 1+\frac{a}{b} \tag{24}$$

通过迭代得到 b/a 的值约为3.6，这相当于 Z_0 约为77Ω。现在得到了所需要的全部信息。

首先，有线电视设备都是以75Ω阻抗为基础的，这是因为它对应于（几乎是）最小的损耗。这里由于功率较低，所以处理功率的能力不成问题。那么为什么这里的标准是75Ω而不是77Ω呢？这是因为工程师们喜欢四舍五入。类似地，这也是采用50Ω的理由（我们终于得到这个数了）。由于77Ω时损耗最小而30Ω时得到最大的功率传输能力，所以一个合理的折中就是取其平均值——这个值经过四舍五入后就得到了50Ω。情况就是如此。

3.5.3　总结

既然我们已经理解了宏观世界是怎样选择了50Ω，那么应当很清楚如果性能不是受限于功率处理能力或互连，我们可以自由地选择非常不同的阻抗值。因此，特别是集成电路的设计工程师有比分立元件设计面临的限制有更大的自由度来选择范围很大的阻抗值。即使是在某些分立元件的电路设计中，也值得考虑采用50Ω（或某些其他标准值）的折中，而不是非要受这些值的限制。

第4章 阻抗匹配

4.1 引言

低频模拟电路的设计者经常对微波工作者对阻抗匹配的执着而感到困惑不解。比如说,模拟电路的书籍几乎没有一章是关于阻抗匹配的。相反,工程师在低频电路的设计中通常用电压增益表明指标,很少或者根本没有提及阻抗匹配。与此截然相反,射频工作者事实上总是非常关注阻抗匹配。他们看问题的不同方法主要是因为在低频的情况下,功率增益是非常容易实现的。比如说,教科书中的变压器能够提供阻抗匹配来实现最大功率增益,但是电子电路的电压放大器比起绕在磁铁上的线圈更容易实现(而且当然是更为方便)。对阻抗变换的需要通常是默认的(如果真是需要的话),而且仅仅是,比如说,用电压缓冲器粗略地变换到随便一个低阻值。但是在另一方面,射频功率增益通常是非常有限的资源,因此我们必须要小心不要去浪费它。阻抗变换网络因此在射频电路中起到非常关键的作用。

除了获得最大功率增益外,还有很多原因要求做阻抗匹配。一个原因是阻抗匹配可以使系统对互连线的长度不敏感。正是因为这个原因,分立的射频和微波器件通常是设计成标准化的阻抗(比如说 50Ω)。这个标准化通过消除它们之间阻抗变换的需要,极大地方便了这些元件组装成一个大的系统。即使这样,连接这些标准元件的输入输出也可能依然需要阻抗变换,如果源或者负载阻抗不等于标准值的话。比如说,这种情况经常在耦合到天线或者其他谐振器时发生。

使功率增益最大化不是要求阻抗匹配的唯一目的。对于每一个放大器,存在一个特定的阻抗值能够使其噪声系数最小。除非是运气好,这个值很少是与最大功率增益匹配或与标准值,例如 50Ω,匹配要求的值一样。这也说明阻抗变换是必须的。

还有一个需要提供匹配的原因是滤波器通常是通信系统中限制性能的元件[①]。滤波器的关键特性,比如说通带和延迟的平坦性,会对驱动和终端的负载十分敏感,因此阻抗匹配再次成为了一个必须的要求。

阻抗匹配在大功率的系统中也是非常有价值的,在那里因为不匹配而引起的驻波可能导致电压的峰值在沿着传输线的不同位置上超过击穿电压。灾难性的击穿可能不是唯一的后果,因为发射机的输出功率和效率很大程度上取决于负载的电阻值。一个不合适的阻抗可能导致效率的下降(因此导致散热问题更加严重),引起工作状态不稳定(包括振荡),使得发射机不能达到指定的输出功率,甚至可能损坏发射机。

为了更仔细地研究阻抗匹配,我们首先给出一个定理,这个定理给我们提供了为什么要

---

① 这个限制在原理上并非是必须的,但工程实践中却经常这么做。因为人们通常总是选择能满足设计目标的最便宜的滤波器。

考虑阻抗匹配的基本道理。

4.2　最大功率传输定理

绝大多数的工科本科生在他们学习过程中的某个阶段都会遇到最大功率传输定理。但是有证据表明，传授给学生有关这个定理的内涵是不够完全与准确的(有关问题是出在老师的讲解不清楚还是学生接受有偏差的争论一直不断；我们在这里也不认为我们能解决这个问题)。这个最大功率传输定理回答的是这样一个问题：给定一个固定的信号源阻抗 Z_S，什么样的负载阻抗能够让传递到负载的功率最大。图 4.1 给出了一个正弦信号驱动的网络②。

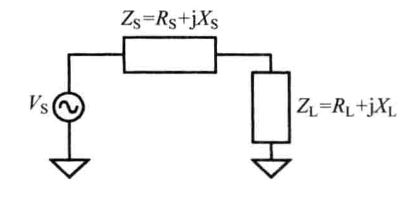


图 4.1　推导最大功率传输定理的网络

信号源传输给负载阻抗的功率是完全由负载电阻 R_L 消耗的，因为电抗成分不会耗费功率。因此，传输给负载的功率就是：

$$\frac{|V_R|^2}{R_L}=\frac{R_L|V_S|^2}{(R_L+R_S)^2+(X_L+X_S)^2} \tag{1}$$

在这里，电压 V_R 和 V_S 分别是负载电阻和电压源两端的均方值(rms)电压。

为了让 R_L 得到的功率最大，很显然 X_L 和 X_S 的代数值应该是相反的(即互为负值)，这样它们加起来为零。也就是说，我们希望纯电抗为零。在一个单频率上或者在一个窄的频率范围内产生这个条件相对来说是容易的。在这个条件($X_L=-X_S$)下，进一步使方程(1)最大化，就得到了 R_L 应该等于 R_S。因此最大功率传输在负载与信号源阻抗共轭匹配的情况下发生。

注意上面这个条件隐指了阻抗变换必须是没有损耗的。比如说，通过加一个串联电阻来保证电压源电阻见到一个相等的电阻并不能使传输到负载的功率最大，因为这个附加的电阻不是我们最终希望输出功率给予的负载的一部分③。

我们因此可以总结最大化功率传输的一般原则：只用没有损耗的元件，消去电抗，同时变换现有的电阻，直到源和负载电阻匹配。

幸运的是，我们有很多种方法来实现这个原则，因为没有一个单一的阻抗匹配的方法是在所有情况下都是理想的。我们因此将提供一些常用的方法，它们或者可以单独使用或者结合起来用。

一些匹配的测试标准范例

所有的工程都是这样或那样的折中选择，阻抗匹配同样是这样。因为完美是不可能达到的，我们必须给阻抗匹配定下实际可达到的目标。然而，没有硬性的和快捷的规则决定什么构成了好的匹配；要求达到的性能跟具体应用十分有关。因此有必要发展几个等价的描述失配的方法，比较常见的有：反射系数 Γ，其等同于 S_{11}；回波损耗(RL)，其简单地就是

---

② 学生经常犯错的地方在于忘记了这个问题叙述中“固定的源阻抗”这部分。如果我们是可以自由选择源阻抗，那么我们应当选择其值为零，而不是等于负载阻抗。然后，我们理论上可以从源中获得无穷大的功率(如果同时选择很小的负载阻抗的话，译者注)。

③ 有时加一个电阻对其他方面是有用的，比如防止放大器的寄生振荡，然而这些和最大功率增益不是一件事。

S_{11}的倒数（当用 dB 度量时，就是其代数相反数，即加负号）；以及驻波比(SWR)，定义为

$$\mathrm{SWR}=\frac{1+|\Gamma|}{1-|\Gamma|} \tag{2}$$

可以回想起 SWR 是线上最高振幅和最低振幅的比值。一个完美的匹配不产生会与前行波交替增强或减弱的反射回来的干扰波，因此得到 1∶1 的 SWR。失配越严重，因建设性增强与破坏性减弱的干扰的振幅不一致性就越严重，因此 SWR 越大。

因为等反射系数幅度的轨迹是简单的同心圆，与史密斯圆图中心的距离是失配测量的直接手段。十分方便的是，所有这些表征量之间的关系可以在一个典型的（且完整的）史密斯圆图上兼顾到。在史密斯圆图的下方有一些辅助规则，能让你从任何一种等价的形式读出测量到的失配。表 4.1 提供了等价的信息。

表 4.1　失配的等价度量形式

$\|\Gamma\|(\|S_{11}\|)$	RL/dB	SWR
0	infinite	1.00
0.025	32.0	1.05
0.05	26.0	1.11
0.075	22.5	1.16
0.1	20.0	1.22
0.15	16.5	1.35
0.2	14.0	1.50
0.25	12.0	1.67
0.3	10.5	1.86
0.35	9.1	2.07
0.4	8.0	2.33
0.45	6.9	2.64
0.5	6.0	3.00

正如我们已经提到的，需要的匹配程度是和应用相关的。尽管这么说，一般说来，大于 3 的 SWR 值很少被许可。一个更典型的这个指标的规定可以表述成最小回波损耗等于 10dB，而超过 15dB 是更为企求的。这些指标分别对应的 SWR 的值大约是 2∶1 和 1.5∶1。

4.3　匹配方法

最大功率传输定理仅告诉我们希望将电抗部分抵消掉，同时将负载的电阻部分变换成等于源电阻。实在有很多方法可以实现这些目标，它们在尺寸、复杂度以及带宽中有不同的折中结果。我们首先看一些相对说来是窄带匹配的技术，然后考虑几个实现宽带阻抗匹配的方法。

4.3.1　经典的集总元件匹配方法

谐振电路回顾

多年来发展起来的极多的窄带阻抗匹配方法可以用一个统一的、基于集总谐振电路的理解来解释。接下来，我们先简要回顾一下 *RLC* 谐振电路的一些特性。

一个并联 RLC 网络在谐振时的 Q 值为

$$Q = R/\sqrt{L/C} \tag{3}$$

其中量 $\sqrt{L/C}$ 具有电阻的量纲，并有时候被称作网络的特征阻抗[④]。它有以下的重要性：它等于谐振时容性和电感性的电抗值。这个可以容易被证明：

$$|Z_C| = |Z_L| = \omega_0 L = L/\sqrt{LC} = \sqrt{L/C} \tag{4}$$

我们会发现这个量时不时会出现[⑤]，所以要记住。

在我们继续下去之前，让我们看看这个有关 Q 的公式是否有对头。当并联电阻趋向于无穷大，那么 Q 值也趋向于无穷大。这个行为看上去是合理的，因为在无穷大电阻的情况下，网络退化成一个纯 LC 系统。当网络里只有电抗元件时，能量是无从被消耗的，所以 Q 应该趋于无穷大，正如上面这个式子表明的那样。另外 Q 值也随着电抗元件的阻抗减小（这是通过减小 L/C 达到的）而增加，因为纯电阻比起电抗性的阻抗来说变得不那么重要了。

为了完整性，我们可以推导几个另外的有关并联 RLC 网络在谐振时的 Q 表达式

$$Q = \frac{R}{|Z_{L,C}|} = \frac{R}{\omega_0 L} = \omega_0 RC \tag{5}$$

这种网络可以实现阻抗转换的能力可以从在谐振时电感和电容支路电流的行为看出。我们能够方便地计算出这些电流会完全不同于流过整个网络的电流（这只简单地是由于并联电阻引起的），从而获得关于阻抗转换技术的一些重要的内在知识。

在谐振的时候，网络上的电压为 $I_{in}R$。因为电感性和电容性的电抗在谐振时是相等的，那么这两个支路的电流幅度也是相等的

$$|I_L| = |I_C| = \frac{|V|}{Z} = \frac{|I_{in}|R}{\omega_0 L} = \frac{|I_{in}|R\sqrt{LC}}{L} = |I_{in}|\frac{R}{\sqrt{L/C}} = Q|I_{in}| \tag{6}$$

也就是说，流过电感和电容性支路的电流是流入整个网络的净电流的 Q 倍。因此，如果 $Q=1000$，我们用 1A 的电流来驱动谐振时的网络，那么这个 1A 的电流会流过电阻，但流过电感和电容的电流将是不得了大的 1000A（直到它们被熔断）。从这个简单的例子里，你可以得到这样一个结论，即仅仅认为电感电容在谐振时互相抵消是个不完整的说法。

我们或许可以从以上的电流的急剧提升中推想到这个网络实现了阻抗下变换的功能。也就是说，在高电流支路的阻抗比总的电路的阻抗要低。为了进一步证实这个想法，我们来考虑图 4.2 的例子。我们选择把一个电阻和电感串联以反映电感一般比电容损耗更大的现实。图中的模型因此是对典型的、在实际上遇到的并联 RLC 电路的一个可行的初步近似。

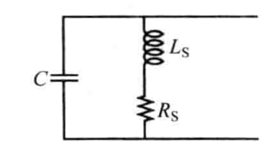


图 4.2　并非纯并联的 RLC 槽路电路

因为我们已经详细分析了纯 RLC 并联网络，我们希望尽可能多地利用那些已经知道的知识。所以，我们通过把图 4.2 的电路中的串联 LR 支路替代成并联支路而将该电路转化为一个纯 RLC 并联电路。显然，这种替换并非通用的，但是在一个有限度的频段内（例如，接近谐振处）这个等价性会是相当合理的。为了正式地证明，我

④　这个项通常应用于传输线中，但是我们可以看到甚至它在集总参数网络里也具备某些重要意义。

⑤　回忆一下传输线的特征阻抗是由同样的表达式给出的，只是那里 L 和 C 是被理解为**单位长度**的电感和电容。

们令串联支路和并联支路的阻抗相等，得到

$$j\omega_0 L_S + R_S = (j\omega_0 L_P) \parallel R_P = \frac{(\omega_0 L_P)^2 R_P + j\omega_0 L_P R_P^2}{R_P^2 + (\omega_0 L_P)^2} \tag{7}$$

如果我们令实部相等，并注意到有 $Q=R_P/\omega_0 L_P=\omega_0 L_S/R_S$⑥，得到

$$R_P = R_S(Q^2+1) \tag{8}$$

类似地，让虚部相等，可以得到

$$L_P = L_S\left(\frac{Q^2+1}{Q^2}\right) \tag{9}$$

我们可以推导类似的一组计算串联和并联 RC 的等价公式

$$R_P = R_S(Q^2+1) \tag{10}$$

$$C_P = C_S\left(\frac{Q^2}{Q^2+1}\right) \tag{11}$$

我们稍停一下来看一下这些变换公式。通过更仔细的观察，可以清楚地看到，我们能得到同时适用于 RC 和 LR 网络的一个统一形式

$$R_P = R_S(Q^2+1) \tag{12}$$

$$X_P = X_S\left(\frac{Q^2+1}{Q^2}\right) \tag{13}$$

这里 X 是阻抗的虚部。用这种方式，我们只需记住这一对通用的公式即可将任何“不纯粹”的 RLC 网络转换为一个纯粹的并联(或串联)的 RLC 网络，这个网络分析起来直截了当。这两个并联项总是比相应的串联项大。我们必须记住这种等价只在中心频率 ω_0 两侧的一个窄的频率范围内有效。

L 形匹配

现在我们以阻抗变换的叙述方式来重新解释所有上面这些结果。正如我们注意到的，谐振 RLC 网络中电压或电流被乘上 Q 暗示它们有阻抗变换的能力。的确，在前面一节里推导出来的串联-并联 RC/LR 网络的变换公式实际上以显式表明了这个性质。为了更清楚地看出来，我们再来观察一下图 4.2 的电路，这一次，稍微变动一下重画为图 4.3。

这里，我们将 R_S 处理为这个网络的负载电阻。当从电容的两端去看这个电阻时，它依照在前一节发展起来的公式被变换成了一个等价的 R_P。仔细看一下那些通用公式，可以明显看出 R_P 总是比 R_S 大，所以图 4.3 所示的网络向上(即向增大的方向)变换了电阻。假若需要一个阻抗下变换，那么我们只需要简单地互换一下端口，如图 4.4 所示。

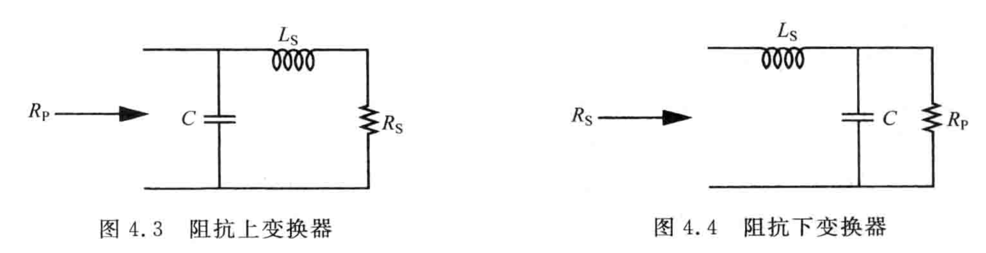


图 4.3　阻抗上变换器　　　　图 4.4　阻抗下变换器

⑥ 假若串联与并联部分是等价的，那么它们的 Q 值必然也是相等的。

有一个很好的直觉方式可跟踪阻抗变换的方向。例如,假若图4.4所示的电路由一个测试电压源驱动,那么结果就是一个并联的 RLC 网络,因为测试电源的戴维宁电阻等于零。现在,这个并联 RLC 网络在谐振时流过电感的电流值等于流过 R_P 电流的 Q 倍。而测试电源所看到的正好是这个增加了的电流,因此等效来说,可以解释为电阻变小了。

同样应当清楚的是电感和电容互换并不改变变换比率,因此我们可以兼顾其他的设计考虑来决定选择高通还是低通电路结构。比如,如果电源或者负载已经具有某些电抗,通过选择合适的拓扑结构有可能将这些不希望的元件吸收到阻抗变换器中去。另一个考虑是匹配的带宽。在这方面,高通的结构式更好一些。这从 Murphy 定理是可以想得到的(Murphy 定理是一个半开玩笑,又半是日常生活中遇到的真理的总结。这个定理说你不想要的事,往往会发生;而想要的事往往是得不到的,译者注):因为低通电路对滤掉不想要的频谱成分经常总是更为有用(因此被经常采用)。

这个电路因为它的电路形状被称为L形匹配(也许你需要躺下倒过来看它才像),而不是因为它包含一个电感。它的主要特点就是简单。然而这个电路只有两个自由度(我们只能选择 L 和 C)。因此一旦确定了阻抗变换比及谐振频率,网络的 Q 值就自动确定了(而 Q 值可能会影响网络的频率选择性质,包括带宽)。如果你希望得到不同的 Q 值,那么就必须采用能提供额外自由度的网络。我们很快将介绍这些网络。

对L匹配的最后一点说明是:如果 $Q^2 \gg 1$,那些通用的公式能够被简化。如果这个不等式能被满足,那么以下的近似公式成立:

$$R_P \approx R_S Q^2 = R_S \left(\frac{1}{\omega_0 R_S C}\right)^2 = \frac{1}{R_S}\frac{L_S}{C} \tag{14}$$

这又可以被重写为

$$R_P R_S \approx \frac{L_S}{C} = Z_0^2 \tag{15}$$

式中,Z_0 是网络的特征阻抗,正如在4.3.1节中讨论的那样。

换一种说法,选择一个匹配网络的特征阻抗等于源阻抗和负载阻抗的几何平均值。当我们考虑经典的四分之一波长传输线变换器时,我们会遇到一个十分类似的解决方法。

我们也可以推导出 Q 近似等于变换率的平方根:

$$Q \approx \sqrt{R_P/R_S} \tag{16}$$

最后,在这个变换过程中,电抗没有太多的变化,即

$$X_P \approx X_S \tag{17}$$

只要 Q 约大于3或4,那么所引起的误差在10%以下。如果 Q 大于10,那么最大误差将在1%左右。至少对于快速的手工草算(或说,用信封的背面这样一小块纸就可算出)来说,这些简化的公式是很合适的。最终的设计值总是可以采用完整的通用公式计算出来。

L形匹配也是一个良好而简单的网络,用来说明史密斯圆图可以用作图形化的阻抗匹配。我们可以把阻抗匹配问题看作是一个游戏,它的目标是将任一个初始位置转回到"家",这个家指的就是史密斯圆图的圆心[7]。游戏的规则是不沿直线运动,这也是这个游戏有挑

⑦ 一个完全等价的描述是去寻找将圆点作为起点到达一个任意给定的终点。同样的运动规则适用:你就是以反向的方式来玩这个游戏,见 http://contact.tm.agilent.com/Agilent/tmo/an-95-1/classes/imatch.html。

战和有意思的地方。实际上，构成史密斯圆图基础的弧形坐标系统意味着我们每步操作都要通过一个圆弧。挑战在于我们如何组合一系列的圆弧使得我们到达圆心。想要设计出我们给出的匹配网络，关键是要熟悉加入并联或串联的元件对轨迹的影响，同样地，你可以去设计你自己新的匹配网络。

加入一个串联的电抗元件简单地就是让我们沿一个恒电阻圆移动。类似地，加入一个并联电纳使我们沿恒电导圆移动。因为阻抗匹配网络通常是由串联和并联元件组合而成，所以使用一个完整的 ZY（阻抗、导纳）史密斯圆图来设计匹配网络对我们是有帮助的。尔后我们在这个混合上增加更多的轨迹，比如对应于加入传输线段及因为频率变化而引起的轨迹。

用一个具体例子可以很好地说明这些要点，所以我们采用史密斯圆图的术语再来审视 L 匹配网络。在图 4.5 里，假定负载电阻 R_S 是如此之低，以至于我们想把它往上变换。在负载上加入一个串联电感可以让我们沿着对应 $R=R_S$ 的恒定电阻圆移动，随后加入一个并联的 C 使之沿恒定电导圆移动。最终我们希望停止在史密斯圆图的中心，因此这个恒定电导圆必须包含 $G=1$ 的终点。为了设计这个匹配，仅是需要画两个圆：一个是对应于 $G=1$ 的恒定电导圆，另一个是对应于 R_S 的恒定电阻圆。这两个圆的重合处就确定了所需要的电感和电容值。接下来只需要逆反归一化即可得到完成设计需要的元件设计值。

正如图 4.6 所示，我们从负载阻抗 R_S 开始加入一个串联电感以使之沿恒定电阻圆顺时针移动。加入的电感值要正好使轨迹到达 $G=1$ 对应的恒定电导圆。这个选择是重要的，因为接下来加入一个适当的电容就可以沿这个圆按顺时针方向转到史密斯圆图的中心。通过读出 P 点对应的并联电纳或者串联电感值，可以容易地分别计算出电容和电感值[⑧]。

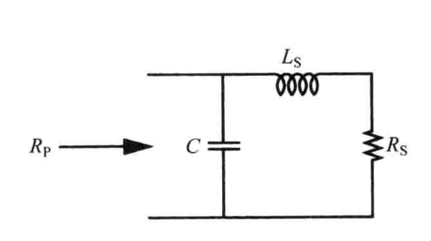


图 4.5　L 形匹配例子

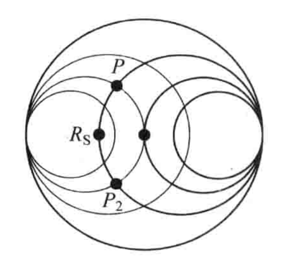


图 4.6　带有 ZY 史密斯圆图的 L 形匹配例子（细线是那些恒电导圆）

注意，这两个圆还有一个交叉点 P_2，这表明还存在一个等效的匹配网络。在这个例子里，一个串联电容使我们沿恒定 R_S 圆按逆时针旋转到 P_2 点，然后加入一个并联电感让我们沿 $G=1$ 的恒电导圆逆时针旋转。加入足够的并联电感以再一次停留在 $G=1$ 的点。P_2 对应的并联电纳和串联电抗让我们可以计算出需要的并联电感和串联电容值，以完成设计。这个对应的网络相当于上变换 L 形匹配的高通版本。

从史密斯圆图的构造，可以看出对复数负载阻抗也容易包括在内。比如，如果负载已经有一个串联电感，那么就不需要加入那么多电感以到达 P 点，接下来的步骤不变。这样，通

⑧　对于每个交汇点，我们只有一个单一的阻抗值。由此，电容和电感值可以被找到。

过将寄生参数视作电路元件的一部分，我们懂得正确地选择匹配网络如何能减小寄生带来的影响。

史密斯圆图的构造也帮助我们记住这种网络拓扑无法实现下变换。如果归一化的负载电阻超过1，那么加一个串联电抗必将让我们沿恒定电阻圆移动，这将永远无法与 $G=1$ 的恒定电导圆相交。那么加入并联元件将无法带我们回到史密斯圆图的中心。这种同样类型的构造过程也让我们推断通过交换端口的位置可以得到想要的阻抗变换。

π 形匹配

正如已经讨论的那样，L 形匹配的一个限制是只能指定中心频率、阻抗变换率以及 Q 值这三个量中的两个。为了得到第三个自由度，可以采用图 4.7 中的网络。

这个电路称为 π 形匹配，它同样是根据形状而得名。理解这个匹配网络的最方便的方法是把它看成是两个 L 形匹配串联在一起，其中一个向下变换，另一个是向上变换，见图 4.8。这里，负载电阻 R_P 先变换成在两个电感连接处的一个较低的电阻(称之为镜像电阻或中间电阻，此处表示为 R_I)，然后这个镜像电阻通过第二个 L 形匹配向上转换为 R_{in} 值。

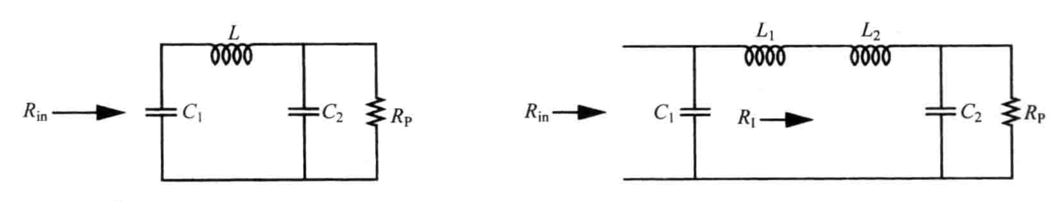


图 4.7　π 形匹配(低通版本)　　图 4.8　π 形匹配看作是两个 L 形匹配的级联

现在看来也许有点像政府部门的项目(指效率低，译者注)，即采用一个 L 形部分向下变换，然后又采用另一个 L 形部分再返回去进行向上变换，然而这样却得到要寻找的额外的自由度。回想一下，对于一个 L 形匹配，Q 值固定在一个大致等于阻抗变换比平方根的值上。典型情况是，一个 L 形匹配的 Q 值不会特别高，因为通常并不需要很大的变换比。π 形匹配通过引入一个变换过去的中间电阻值使 Q 值与变换比分开来，这使得我们即使在整个变换比不是特别大的情况下也能得到比在 L 形匹配中得到的高得多的 Q 值。

现在我们有了三个自由度(两个电容，以及两个电感的和)，因此可以独立地指定中心频率、Q 值(或带宽)以及整个阻抗变换比。然而，与 L 形(或任何其他种匹配)一样，有可能出现不切实际或不很方便的元件数值，因此也许需要有某些创造性或折中考虑来实现一个合理的设计。在许多情形中，把几个匹配网络级联起来也许会有帮助(例如，用更多的电路段来扩展 π 形匹配的概念)。

最后需要注意的一点：π 形匹配被比较广泛使用的另一个原因是任何连到网络的寄生电容都可以被吸收到网络设计中去。这个性质很有价值，因为在实际例子经常遇到的主导寄生元件是电容。

T 形匹配

π 形匹配是以一种特定方式通过串联两个 L 形部分得到的。把 L 形部分连成另一种形式就得到了 π 形匹配的对仗电路——T 形匹配，如图 4.9 所示。这里本来在实现时是单个的电容现在已经被明确地分成了两个分开电容，以此来清楚地表明这个网络是由两个通常的 L 形匹配在它们的高阻抗端口连接起来的。这个(并行的)镜像电阻是从这两个电容的

两端看上去的，就像在 π 形匹配中那样，或者向右看或者向左看。

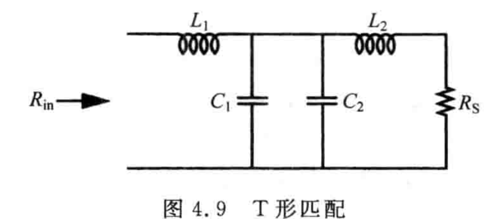


图 4.9　T 形匹配

T 形匹配在性能上和 π 形匹配是完全等效的。当源和负载端的寄生参数主要是电感性质的，可以作为网络的一部分时，这是一个较好的选择，就像在 π 形匹配中端口电容可以被吸收入网络一样，它们也可以被吸收进网络中。

注意，也存在着其他形式的 L 电路段的级联方式。这些方式在不经常遇到的场合下值得考虑。但是因为它们实在是少见，我们这里只是提一下它们的存在。

抽头电容和电感

电阻分压器可以变换阻抗但是损耗很大。如果我们将电阻换为纯电抗，那么形成的网络仍然可以变换阻抗，但不会再有损耗。这时还有剩余的电抗，但是可以另外调谐掉。这种电路的一个例子如图 4.10 所示。

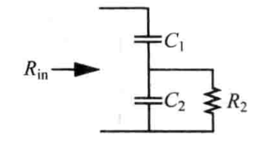


图 4.10　带抽头的电容匹配网络

这个阻抗变换器的工作可以最好地被理解为一个电容分压器。在一个完全没有损耗的网络中如果要保持功率不变，那么减小电压必须同时使阻抗按照电压减小的平方成正比地减小。虽然这整个网络并不是完全无损耗的，但我们仍希望阻抗变换比大致为

$$\frac{R_2}{R_{in}} \approx \left(\frac{1/sC_2}{1/sC_1 + 1/sC_2}\right)^2 = \left(\frac{C_1}{C_1 + C_2}\right)^2 \tag{18}$$

这样，这个网络或者把电阻 R_{in} 向下变换成 R_2 值，或者说把电阻 R_2 向上变换成 R_{in} 值。

为了确认这一点，单独分析一下这个有电阻负载的电容分压器。稍加推导就可以容易地求出这一组合的导纳

$$Y_{in} = \frac{j\omega C_1 - \omega^2 R_2 C_1 C_2}{j\omega R_2 (C_1 + C_2) + 1} \tag{19}$$

它的实部部分是

$$G_{in} = \frac{\omega^2 R_2 C_1^2}{\omega^2 R_2^2 (C_1 + C_2)^2 + 1} \tag{20}$$

在足够高的频率时，等效的并联电导的确可以被简化为

$$G_{in} \approx \frac{\omega^2 R_2 C_1^2}{\omega^2 R_2^2 (C_1 + C_2)^2} = G_2 \cdot \left[\frac{C_1}{C_1 + C_2}\right]^2 = \frac{G_2}{n^2} \tag{21}$$

这正如我们预期的一样。公式(21)也定义了一个系数 n，它是一个能产生与电容分压器具有相同变换比的理想变压器的匝数比。在统一处理各种振荡器的分析时，这个等效匝数比的概念将被证明特别有用。

为完整起见，还可以计算出这个导纳的虚数部分：

$$B_{in} = \frac{\omega C_1 + \omega^3 R_2^2 C_1 C_2 (C_1 + C_2)}{\omega^2 R_2^2 (C_1 + C_2)^2 + 1} \tag{22}$$

上式在足够高的频率时接近如下的极限值：

$$B_{in} \approx \omega \cdot \frac{C_1 C_2}{C_1 + C_2} = \omega \cdot C_{eq} \tag{23}$$

这里电容 C_{eq} 等效于两个电容串联。

前面的一系列公式能很好地用于分析，尤其是在建立设计的直觉上。式(21)和式(23)对于初次粗略的设计也是极为有用的。

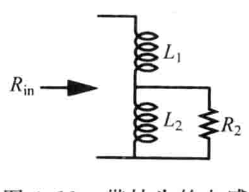


图 4.11 带抽头的电感变换器

现在简略考虑一下抽头电感作为匹配网络(见图 4.11)。正如你可以想到的，它的特性非常类似于相应的抽头电容匹配网络。我们不再进行设计公式的详细推导，这完全类似于抽头电容的情形，但可以观察到：R_2 必定小于 R_{in}，因为这里仍然是一个分压器。

显然，抽头作为阻抗变换方法的应用比它在纯电容或纯电感网络中的应用要广得多。的确，我们将会看到耦合到一个谐振器或从其上耦合出来，通常是通过使用阻抗变换抽头结构来完成的。

4.3.2 经典的传输线阻抗变换器

当频率增加时，集总元件的实现方法逐渐变得不合适，因此我们需要扩展选择范围来包括分布式的匹配方法。

迄今为止，最著名的是 $\lambda/4$ 传输线变换器。回想起采用史密斯圆图，在负载上加入一段传输线将使我们沿一个圆弧(又是一个圆弧)行进，这一次这个圆弧是以原点($R=1$)为圆心的等反射系数圆。再回想起，每经过 $\lambda/2$ 的线，阻抗将会重复，这在史密斯圆图上代表的是一个完整的圆。因此，加入一段 $\lambda/4$ 的传输线将绕等反射系数圆转半圈。所以，给定一个负载阻抗 Z_1，那么在线的另一端将看到它的倒数(要计及一个归一化常数)。具体的是

$$Z_2 = Z_0^2 / Z_1 \tag{24}$$

这种关系告诉我们，我们可以使用 $\lambda/4$ 传输线通过选择合适的特征阻抗来做阻抗变换器。这个特征阻抗是

$$Z_0 = \sqrt{Z_1 Z_2} \tag{25}$$

这种变换器实际上是窄带的，因为它的长度只在一个频率处满足 $\lambda/4$ 标准。另一个需要考虑的是实际可实现的阻抗是受到限制的，尤其对于微带线的形式。这样，实现一个 10～15Ω 或者 150～200Ω 的微带线变换器是比较困难的。

如果(比如在低频下)需要的线的长度不切实际地大或者需要的特征阻抗不合适，那么集总元件的实现是所希望的。最简单的是用 π 形或者 T 形网络去近似 $\lambda/4$ 传输线，见图 4.12。

图 4.12 采用集总元件的 $\lambda/4$ 阻抗匹配

这些参数值由以下简单关系给出：

$$L = \sqrt{R_1 R_2} / \omega_0 \tag{26}$$

$$C = 1/\omega_0 \sqrt{R_1 R_2} \tag{27}$$

这里 R_1 和 R_2 分别是源和负载电阻。可以看到这些公式使电感和电容的电抗值等于源和

负载阻抗的几何平均值。

读者可能注意到这种网络其实是 π 形匹配的一个特例。它的特点是容易设计。

$\lambda/4$ 和 $\lambda/2$ 的传输线有着众所周知的阻抗变换特性：前者是倒数关系，后者是重复。因为复数值的负载仍然比较复杂，完成一个匹配需要额外的元件以抵消剩余的虚部。

还有一个不是太常见的线长，它的特性在阻抗匹配的设计中也很有用。$\lambda/8$ 的传输线能将复数值阻抗变换为纯实的。为了演示这一特性，我们首先将终端接任意负载的无损传输线的输入阻抗表示：

$$Z(z) = Z_0 \frac{Z_L + jZ_0 \tan\beta z}{Z_0 + jZ_L \tan\beta z} \tag{28}$$

这里负载所在的坐标 $z=0$，其他地方取正值。设 βz 等于 $\pi/4$ 可以得到

$$Z = Z_0 \frac{Z_L + jZ_0}{Z_0 + jZ_L} = Z_0 \frac{(R_L + jX_L) + jR_0}{R_0 + j(R_L + jX_L)} = Z_0 \frac{R_L + j(R_0 + X_L)}{(R_0 - X_L) + jR_L} \tag{29}$$

将分母有理化，求解得到

$$Z = Z_0 \frac{R_L(R_0 - X_L) + R_L(R_0 + X_L) + j[(R_0 + X_L)(R_0 - X_L) - R_L^2]}{(R_0 - X_L)^2 + R_L^2} \tag{30}$$

很明显，阻抗并不是自动变为纯实部的，但是很容易确定虚部为 0 的条件：

$$(R_0 + X_L)(R_0 - X_L) - R_L^2 = 0 \Rightarrow R_0^2 = X_L^2 + R_L^2 \tag{31}$$

这样，如果我们选择 $\lambda/8$ 波长传输线的特征阻抗等于负载阻抗的幅度，那么变换后的阻抗将变为纯实部，它的值为

$$Z = Z_0 \frac{R_L(R_0 - X_L) + R_L(R_0 + X_L)}{(R_0 - X_L)^2 + R_L^2} \tag{32}$$

经过简化，变为

$$Z = Z_0 \frac{R_L}{R_0 - X_L} = Z_0 \frac{R_L}{|Z_L| - X_L} \tag{33}$$

注意到这些变量之间的关系保证了非负的电阻，即保证了无源性。另外注意到，当负载接近一个纯电抗，变换后的电阻接近无穷大。又一次我们看到，仅根据能量守恒的原则就可以预计到这个结果：如果一个无损的传输线加上一个无损的电抗不会导致一个有损的阻抗。

$\lambda/8$ 传输线尽管可能有用，但很少能靠它就可以实现阻抗匹配，因为仅是把复阻抗变换为实部是不够的。最终的目标不是产生一个任意的实阻抗而是变成 Z_0。一个经典的更通用的微波匹配方法是单断线匹配(图 4.13)。它利用了非匹配负载的传输线的阻抗变化，其工作原理可以从史密斯圆图的轨迹上看出。加入一段线使阻抗沿等反射系数圆移动。因为我们将加入一并联阻抗(短断线)，所以我们希望这个断线的加入位置与负载的距离足够长把我们带到 $G=1$ 对应的等电导圆。然后可以加入足够长的短断线使我们转到史密斯圆图的中心。

正如图 4.14 看到的，加入一段线将让我们沿着负载 Z_L 对应的反射系数圆逆时针旋转。有两个线长可以使我们到达想要的恒定电导圆($G=1$)，通常可以选用两个线长中短的一个，就像这里所选的那样，但这并不是必须的[9]。我们选择短的是因为实际线是有损耗的，

---

⑨　这种选择不仅仅是减小了所需要的传输线的长度，它同时经常导致更宽一点的宽带匹配。作为一个规则，通常选择一个最短长度的方案是一个好主意，除非的确存在有某些其他更重要的考虑(例如，偏置电路的便利实现)。损耗是更小了，辐射的机会通常也消失了，而且整个结构也占用更少的面积。

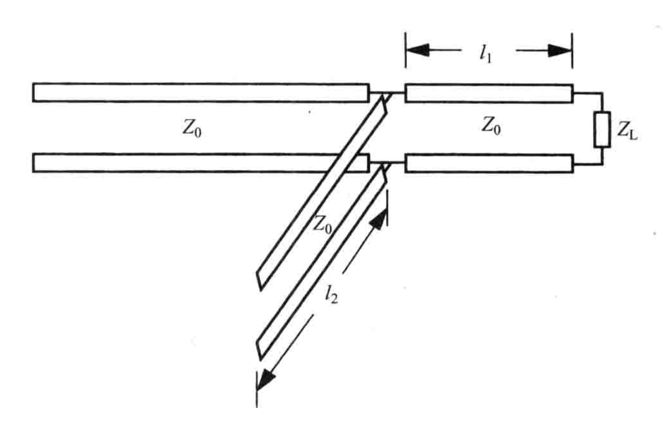


图 4.13　带有单短断线的匹配示意图

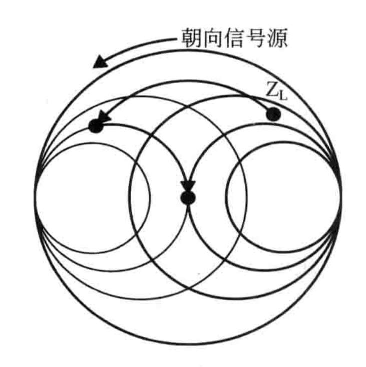


图 4.14　在史密斯圆图上带有单短断线匹配的例子(细线是导纳圆)

这样可以减小损耗,结构紧凑。然而在某些条件下(比如在很高频率下),需要的线长可能非常短。在这些情况下,选择长一点的线,甚至加入整数倍半波长的线可能更合适。其他的实际考虑,比如说如何容易给放大器提供偏置,也可能导致另外一种选择。

一旦线长 l_1 选定,所剩下的就是选择合适的并联短断线以抵消剩下的阻抗虚部。理论上来说,一个终端为短路和开路的线都可以完成所需要的阻抗。一般文献里推荐使用短路短断线基于以下几个原因。一个是近似一个好的短路比一个好的开路更容易一些,这主要是因为边缘电场将使开路效果变差。另一个是开路导致的辐射可能干扰短断线的阻抗,更不要说有可能产生和受到干扰。与这些好的方面相悖的是用微带线实现短路是不方便的。经验表明使用开路短断线一般来说尽管有上述种种原因仍然会产生好的结果,因此我们假设完成一个阻抗匹配将使用一段合适长度的开路线。在我们的例子里,注意到选择一个短的线已经带我们到了等电导圆上了,现在只需加一个并联电容即可完成匹配。如果采用其他的线的长度,将需要一个电感性的短断线。

从完整性考虑,需要注意到在理论上可以使用串联短断线来完成阻抗匹配。然而这个方法比并联方法更麻烦,因为它需要将原来的连接打断。因此对于微带线来说,并联短断线是首选的。

单短断线调谐器尽管非常有效,依然需要并联短断线连接至离负载一定的距离上。如果负载阻抗或距离不能精确确定(或者由于制造过程中产生的误差),那么我们必须作调整。遗憾的是,提供一个可调位置的短断线不是一件容易的事,就像设计短断线长度可调一样。商用的调谐器一般是同轴线,并经常使用可滑动的短路线(通俗语称长号),但这种滑动的设计对微带线来说非常不方便。

适应负载阻抗的大的变化不是一件太困难的事。多重调谐器包含由固定距离隔开的两个或三个(或偶尔更多)单短断线调谐器,见图 4.15。由这个额外的短断线提供的额外的设

计自由度可以部分地抵消因固定短断线长度而损失了的自由度，使得双短断线的设计有着比较宽的调谐范围。在很多可能的选项中，最常见的是间距为 $\lambda/8$ 的多重调谐器。我们可以证明，使用这个距离，谐振器可以对其导纳的电导部分小于 $2G_0$ 的任意负载（这个负载是指从图 4.15 中 X 点向右边看进去负载）提供匹配。尽管我们在这里画的是开路短断线，但实际上经常用可滑动的短路线。但是要再指出的是，由于这种短路型的谐振器不方便设计，很难在微带线形式中看到这种结构。

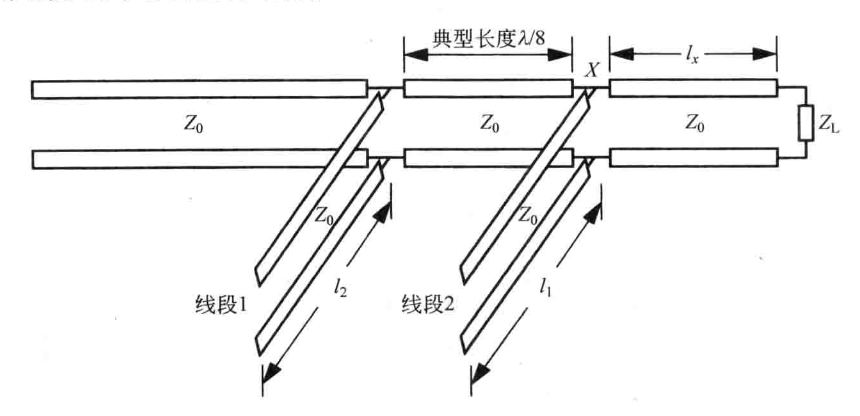


图 4.15　双短断线调谐器（短断线的终端是开路的，图中没有给出实现长度变化的机构）

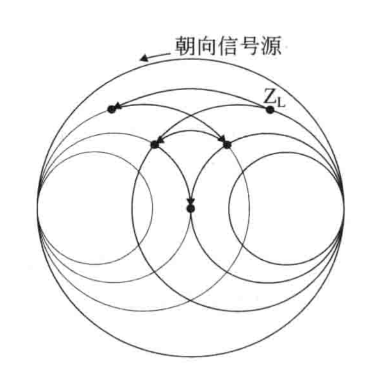


图 4.16　在史密斯圆图上的双短断线阻抗匹配的例子（细线是导纳圆）

可以从图 4.16 看出，史密斯圆图对这种谐振器的描绘可能有些复杂。然而仍然可以找出一个圆弧的组合来构成这个匹配问题的解。在这个例子里，通过一系列拐弯形的圆弧到达原点。和其他例子一样，我们看到的这些轨迹在可行方案中并不是唯一。

第一段圆弧对应于负载和短截线 1 之间线长为 l_x 的转换。习惯上把谐振器和负载放置近一些，使得 l_x 比一个波长短得多，但这种选择并不是绝对的。在图 4.16 中随意选择了一个值 l_x。从负载到第一个谐振器之间的圆弧是在一个等反射系数圆上，它体现了负载和传输线之间的失配。

第二个圆弧对应于短断线 1 的变换。这里假设这个断线的长度足够短这样终端开路的断线将等效一个电容性的电纳。因此，圆弧将沿等电导圆顺时针转动。

第三个圆弧再一次带着我们沿等反射系数圆移动。假设短截线 1 的长度已经被调整到使第三个圆弧可以终结在 $G=1$ 的等电导圆上。从这点出发，短断线 2 的电容带来的电纳使导纳沿 $G=1$ 的圆顺时针移动以最终实现匹配。

我们看到双短断线谐振器提供了使导纳中的实部分步以螺旋式的方式前进到 $G=1$ 的圆上。在实践中，这种调谐一般是通过试探-纠正的方法，它比从理论上推导并计算的方法快得多。

4.3.3 宽带阻抗匹配技术

迄今为止讲述的匹配方法都是基于窄带技术的，因为设计过程是在单一的频率点寻求恰当的阻抗转换。因此我们必须接受在工作频率带宽内阻抗的变化。幸运的是，匹配在比较小的带宽里通常是可以接受的。如果要在更宽的频率范围内实现好的匹配，则必须采用更先进的设计方法，导致转换网络变得更复杂。在描述这些方法前，先看一个由 Hendrik Bode 首先推导出来、后被 Robert Fano 加以扩展的重要的限界公式：[⑩]

$$\int_0^\infty \ln\left[\frac{1}{|\Gamma(\omega)|}\right]d\omega \leqslant \frac{\pi}{\tau} \tag{34}$$

这里 τ 是 Bode 最初推导的简单的单极点负载阻抗的时间常数(Fano 将这个分析扩展至任意负载)。这个不等式告诉我们在非零带宽内实现完美匹配是不可能的(最好情况是我们可以在一个有限频率区间内某几个频率点处实现完美的匹配)。这就是说在设计一个给定频率段内的最佳匹配时是以在所有其他频率处最大可能的失配为代价的。这种方法和设计滤波器的理念是相同的，我们希望在某些频率处实现近乎于 0 的传播损耗而在其他频率处有无限大的衰减。因此，适用于传统滤波器的设计概念也适用于这里。比如，我们可能想到将负载吸收到滤波器中，以在某个通带内产生或逼近切比雪夫型的反射系数行为。我们将在单独的章节讨论滤波器的设计，但现在需要认识到宽带的阻抗匹配问题实际上和带通滤波器的设计问题是等价的。

可以将式(34)变换成更有启示性的形式以对于某个给定反射系数提供匹配的上限带宽(同样，假设是单极点负载)。注意到反射系数为单位 1 时是对积分没有贡献的。如果假设一个(无法达到的)理想的情况，即在某个带宽为 $\Delta\omega$ 内反射系数为一个均匀的值 Γ_{match}，而在带宽外值为单位 1，这时可以计算积分为

$$\int_0^\infty \ln\left|\frac{1}{|\Gamma(\omega)|}\right|d\omega = -\int_0^{\Delta\omega}\ln[|\Gamma_{match}|]d\omega = -(\Delta\omega)\ln[|\Gamma_{match}|] \leqslant \frac{\pi}{\tau} \tag{35}$$

解 Γ_{match} 这个式子可以得到：

$$|\Gamma_{match}| \leqslant \exp\left[-\frac{\pi}{(\Delta\omega)\tau}\right] \tag{36}$$

通过式(35)解得匹配带宽的表达式也是有帮助的：

$$\Delta\omega \leqslant -\frac{\pi}{\tau\ln[|\Gamma_{match}|]} \tag{37}$$

或者将带宽用赫兹为单位表示：

$$\Delta f \leqslant -\frac{1}{2\tau\ln[|\Gamma_{match}|]} \tag{38}$$

因此可以看到允许的非匹配度越大，那么可以达到这个匹配度的带宽也就越大。为了提供一个定量的概念，允许在通带内的 Γ_{match} 为 0.2(对应于 1.5 VSWR)就可以被接受，这样意味着匹配的带宽不会超过 $0.31/\tau$ Hz。在这种情况下，匹配带宽约等于两倍的 −3dB 负载

---

⑩ H. W. Bode, *Network Analysis and Feedback Amplifier Design*(《网路分析与反馈放大器设计》), Van Nostrand, New York, 1945。Robert M. Fano 在他的科学博士论文中扩展了 Bode 的在匹配理论上的工作。这篇博士论文的摘要可以在"Theoretical Limitations on the Broadband Matching of Arbitrary Impedances"(对任意阻抗的宽带匹配的理论极限)*J. Franklin Institute*, January 1950, pp. 57-83 (part Ⅰ) and February 1950, pp. 139-154 (Part Ⅱ)。

带宽。

一个很有用的宽带匹配元件是一个由多节分段特征阻抗值组成的传输线，如果阻抗的变化是连续的，那就是一个缓变阻抗(tapered)传输线。尽管它们只限于纯实部负载(即电阻)的宽带转换，这种变换器还是非常有用的。这种设计的物理的直觉动机十分简单：如果我们需要将两个不同阻抗的线连在一起(因此可见尺度也是不同的)，那么在这两个传输线之间使用足够平滑渐变与连续的过渡可以避免大的反射的产生。

将这个思想定量化，看单节的 $\lambda/4$ 变换器在中心频率外是如何工作的。我们先看

$$Z(z) = Z_m \frac{Z_L + jZ_m \tan\beta z}{Z_m + jZ_L \tan\beta z} \tag{39}$$

这里 Z_m 是匹配节的特征阻抗。

当传输线精确为 $\lambda/4$ 时，可以得到熟悉的阻抗变换式：

$$Z(z) = \frac{Z_m^2}{Z_L} \tag{40}$$

当用作阻抗匹配器件时，将 $\lambda/4$ 传输线的特征阻抗设定为源阻抗和负载阻抗的几何平均值：

$$Z_m = \sqrt{Z_0 Z_L} \tag{41}$$

这是理想情况，现在考虑当频率偏高或偏低于正常值时会发生什么。采用一个最有用的方式，即首先将阻抗关系用反射系数表示：

$$\Gamma = \frac{\dfrac{Z(z)}{Z_0} - 1}{\dfrac{Z(z)}{Z_0} + 1} = \frac{Z_m(Z_L + jZ_m\tan\beta z) - Z_0(Z_m + jZ_L\tan\beta z)}{Z_m(Z_L + jZ_m\tan\beta z) + Z_0(Z_m + jZ_L\tan\beta z)} \tag{42}$$

通过合并项，简化然后取其幅度得到

$$|\Gamma| = \left|\frac{Z_L - Z_0}{Z_L + Z_0 + j2\sqrt{Z_0 Z_L}\tan\beta z}\right| = \frac{|Z_L - Z_0|}{\sqrt{(Z_L + Z_0)^2 + 4Z_0 Z_L(\tan\beta z)^2}} \tag{43}$$

现在，当频率对应于 $\lambda/4$ 时，正切值非常大。同时在此处可以将正弦值近似为 1，这样可以在小的邻域内将反射系数近似为

$$\begin{aligned} |\Gamma| &\approx \frac{|Z_L - Z_0|}{\sqrt{4Z_0 Z_L(\tan\beta z)^2}} = \frac{|Z_L - Z_0|}{2\sqrt{Z_0 Z_L}\,|\tan\beta z|} \\ &\approx \frac{|Z_L - Z_0|}{2\sqrt{Z_0 Z_L}}\,|\cos\beta z| = \frac{\left|\dfrac{Z_L}{Z_0} - 1\right|}{2\sqrt{\dfrac{Z_L}{Z_0}}}\,|\cos\beta z| \end{aligned} \tag{44}$$

因为我们主要关心的是当频率略微偏移中心点时反射系数的变化，所以将反射系数显式地表示为偏移量的函数：

$$|\Gamma| \approx \frac{\left|\dfrac{Z_L}{Z_0} - 1\right|}{2\sqrt{\dfrac{Z_L}{Z_0}}}\,|\cos\beta z| = \frac{\left|\dfrac{Z_L}{Z_0} - 1\right|}{2\sqrt{\dfrac{Z_L}{Z_0}}}\,|\sin\theta| \tag{45}$$

这里 θ 包含相对中心频率的偏移量信息，以相位角的形式表示：

$$\theta = \pi/2 - \beta z \tag{46}$$

从式(45)可以看出，对于中心频率固定的偏移量，当阻抗变换比率提高时，反射系数的幅度变得更差(即更大)。对于非常大的失配，增长率正比于阻抗比率的平方根。这样，我们要求的 $\lambda/4$ 的阻抗变换比率越大，那么在偏移中心频率处的失配越明显。对于很小的 θ，增长近似是线性的(这个可以根据正弦函数的小值近似推导出来)，当角度变大时增长将会变得更快(大约 θ^3)。

如果 Γ_{max} 是最大容许的反射系数，那么可以直接解公式(45)得到相位偏差界限：

$$|\Gamma_{max}| \approx \frac{\left|\frac{Z_L}{Z_0} - 1\right|}{2\sqrt{\frac{Z_L}{Z_0}}} |\sin\theta_{max}| \Rightarrow |\theta_{max}| \approx \arcsin\left[\frac{2|\Gamma_{max}|\sqrt{\frac{Z_L}{Z_0}}}{\left|\frac{Z_L}{Z_0} - 1\right|}\right] \tag{47}$$

下一步是将频率显式引入这个表达式中以计算反射系数幅度在某个最大值下所对应的带宽。

首先，回想起相位常数 β 表示为

$$\beta = \omega/\upsilon_p \tag{48}$$

下一步，我们知道传输线的物理长度对应于中心频率 ω_0 为 $\lambda/4$：

$$\theta = \frac{\pi}{2} - \beta z = \frac{\pi}{2} - \frac{\omega}{\upsilon_p}\frac{\lambda}{4} = \frac{\pi}{2} - \frac{\omega}{\upsilon_p}\frac{\upsilon_p}{4f_0} = \frac{\pi}{2}\left(1 - \frac{\omega}{\omega_0}\right) \tag{49}$$

我们对 $\pm|\theta_{max}|$ 对应的频率边界感兴趣。如果将这两个频率称为 ω_1 和 ω_2，通过简单的运算可以得到

$$\left(\frac{\omega_2 - \omega_1}{\omega_0}\right) = \frac{4}{\pi}|\theta_{max}| \approx \frac{4}{\pi}\arcsin\left[\frac{2|\Gamma_{max}|\sqrt{\frac{Z_L}{Z_0}}}{\left|\frac{Z_L}{Z_0} - 1\right|}\right] \tag{50}$$

公式(50)因此是给定反射系数容差和阻抗变换比下的一个分数带宽的近似(乘上 100 就得到百分比)。例如，我们假定容许的反射系数幅度为 0.2(对应于实际的 SWR 为 1.5)，阻抗变换比为 4. 在这些条件下，归一化的相对带宽为 0.344。如果将目标定得更严格些，即最大反射幅度为之前的一半，0.1(SWR＝1.22)，这样带宽就减半了，从 34％ 变到 17％。

来补充之前所有的数学推导，有必要定性地理解为何当阻抗变换比增大时匹配的频率敏感性会变坏。首先，我们要意识到能得到匹配本身就是一个奇迹，因为我们有两个界面(在变换线的两头)，在那里阻抗失匹是存在的。尽管在这些接口处会有无法避免的反射，$\lambda/4$ 变换器依然可以在一些离散的频率点上实现阻抗匹配。这种变换器不是通过在接口处不让反射发生而产生以上的结果，而是它设法将不可避免的反射加以抵消掉。端口处的阻抗失配越严重，所要求的抵消越难做到。因此问题不是"为何匹配变差?"，而是"它怎么会是匹配好的?" 这样，在频率偏离中心点时会发生失配就不奇怪了，而且在界面上的失配越严重，频率的通带也越窄。

知道了这个内在发生的细节，可以通过减小界面失配本身来减弱失配的加剧，因此也减少了对这个神奇的反射抵消的依赖。但是，一个较小的阻抗失配意味着比较低的阻抗变换比。为了解决这个问题，我们可以简单地，比如说用两段 $\lambda/4$ 传输线级联起来，以提供总的

阻抗变换比，这样每一段可以只承担总负荷中的一部分。很明显，可以将这个过程扩展到任意数量的段数。源于这种思想的宽带阻抗匹配器件因此被称作（多段）步进阻抗变换器。段数越多，带宽越大，尽管是以总的长度增大为代价的。问题自然就产生了：需要多少段和如何选取各段的阻抗以实现设计目标。就像一个人会想到的，某些阻抗的分布可能会更好些。正如滤波器设计（我们将单独讨论这个课题），我们可以探求一系列的通带特性，比如最大平坦度或等波纹的反射系数等，总之是在性能和设计复杂度之间进行折中。

更严格地探索上面这些想法会牵涉到相当复杂的数学，反而会使设计直觉模糊化。因此，我们不会进行严格的推导。然而我们将采用的近似不会过于理想以至于得到的结果与实际相差太大。要求高的读者可以不做简化假设，来进行独立的推导。

考虑一段线的两个界面以及由于阻抗失配带来的反射（见图 4.17）。入射波在 Z_1-Z_2 接口处会经历部分的反射。没有反射的部分继续走到 Z_2-Z_3 接口，发生另一个反射。此时的反射波在 Z_2-Z_1 端口经历再一次的部分反射。而没有反射的部分继续前行和输入波的反射部分叠加，而反射部分回头又到 Z_2-Z_3 接口，照此反复继续。如果忽略所有其他项（尤其是涉及反射系数相乘的项），只保留两个最基本的反射，得到：

$$\Gamma_{\text{tot}} \approx \Gamma_1 + \Gamma_2 \exp(-\mathrm{j}2\theta) \tag{51}$$

这里 θ 是线的电学长度以相位角度表示，而 Γ_1 和 Γ_2 分别是 Z_1-Z_2 接口和 Z_2-Z_3 接口处的反射系数。这个指数项仅是表明这样一个事实，即从第一个界面发出的波传到第二个界面，又反射回到线的输入端口所来回传播用去的 2θ 延时。

读者可以验证这个估算总的反射系数的公式对于 $\lambda/4$ 传输线变换器可以产生正确的结果（即 0）。

现在使用等式(51)，这是一个引用小反射系数近似的结果，作为计算单元去构建由多节级联而成的更大的结构（因为忽略了多重反射我们隐含地假设反射系数的乘积相对于 1 来说很小），参看图 4.18。为了简单起见，假设所有节都有同样长度（即相称的），电学长度为 θ。这种选择是为了简化接下来的推导，但这并不意味着是最优的选择。

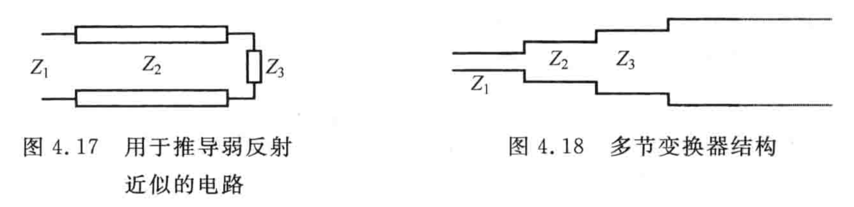


图 4.17　用于推导弱反射近似的电路

图 4.18　多节变换器结构

现在用弱反射近似来一段段地得到：

$$\Gamma_{\text{tot}} \approx \Gamma_1 + \Gamma_2 \exp(-\mathrm{j}2\theta) + \Gamma_3 \exp(-\mathrm{j}4\theta) + \cdots \tag{52}$$

这里各种 Γ_n 是在不连续点的反射系数。如果在假设电学长度相等的线段的基础上，再假设最终的负载只有实部，并且所有的阻抗从结构的一端到另一端单调地变化，那么所有的反射系数都是正实数。再一次强调，这些假设并不是必须被满足，但是用了这些假设使得我们可以获得十分有用的有关设计的内在见解（而且也符合通常的工程实践）[11]。

[11] 比如说，放宽单调变化的条件会允许某一类滤波器的设计（尽管弱反射近似通常不足以提供好的分析结果）。

仔细看一下公式(52)使人联想到傅里叶级数。实际上,它就是一个傅里叶级数。认识到这种序列可以表达具有一般特性的函数,可以想出一种综合方法,即选择每个反射系数以逼近期望的反射系数[12]。取决于具体优化的目标,可以选择反射系数来使得在某个给定的通带内最大反射系数的最小化或者确保单调性。再一次,采用近似的基本动机是与指导滤波器设计的方法是一样的,将稍后在另一章中专门讨论。类似地,也可以选择反射系数以产生最大平坦化响应(所谓的巴特沃兹(Butterworth)条件。这个条件是说在中心频率处反射系数的导数为零的个数最多)或者比如说具有等纹波的特征(就如在切比雪夫逼近下,在那里波纹大小被限制在某个界限内)[13]。

我们已经看到步进式阻抗变换器的性能会随着节数的增加而改善,这是由于每一节的变换比减小了。在它的基础上稍作变动便可得到连续阻抗型变换器。正如有很多选择来决定多节变换器中的阻抗步长,也有许多文献是有关如何在渐变匹配器件中阻抗选择的时机。到了第二次世界大战时,工程师们已经将渐变的声学喇叭(其实现了一种完全类似阻抗变换的功能)转换成电磁形式[14]。特别是指数型渐变传输线有着持久广泛的应用,因为它们比较容易设计与理解。正如其名字所指,这种传输线的阻抗与位置有着简单的指数依赖关系:

$$Z(z) = Z_1 \exp(kz) \tag{53}$$

如果线的总长度为 L,并且要将阻抗 Z_1 变换到 Z_2,那么渐变常数 k 立即就可以由下式决定

$$k = \frac{1}{L}\ln\frac{Z_2}{Z_1} \tag{54}$$

为了实现一个实际的设计,我们需要在总长度选择上的指导原则。直觉上讲,看上去更加缓变的渐变传输线可以实现更好的匹配(或者说更大的 L,或更小的 k),实际情况的确也是如此。一个有用的准则是选择总长度最小为需要进行阻抗变换的最低频率处对应波长的一半。这样,

$$k \leqslant \frac{2}{\lambda_{\max}}\ln\frac{Z_2}{Z_1} \tag{55}$$

一旦渐变常数和线长度已经选定,阻抗随位置变化的函数就已经完全确定了。剩下的就是将有关阻抗的信息转换为物理的尺寸以构建实际的传输线。然而,又一次我们必须考虑实际可实现的阻抗的限制。正如我们提到过的,一般微带线的阻抗是局限在 50Ω 的 3 倍或 4 倍(或大或小)因子之间。

第二次世界大战后十年进行的后续工作表明指数渐变型并不是电学上最优的(尽管你

---

[12] 现在我们可以认识到选择线段等长的好处:它直接导致了反射系数函数看上去像一个傅里叶级数。然而,这个选择并不是必然是最优的,我们鼓励读者去考虑其他的近似方法来允许线段的长度也可以有分布,就如同线段的阻抗可以有分布那样。

[13] 产生巴特沃兹的最大平坦化条件的步进变换器有时也被称为二项式变换器,因为其系数是以二项式展开的形式给出的。

[14] Harold A. Wheeler, "Transmission Lines with Exponential Taper"(具有指数渐变关系的传输线), *Proc. IRE*, January 1939, pp. 65-71。同样,可以参看 George L. Ragan, *Microwave Transmission Circuits*《微波传输线电路》(MIT Rad. Lab. Ser., vol. 9)McGraw-Hill, New York, 1948。关于声学喇叭,我们要感谢瑞利(Rayleigh)爵士。他在声学上的工作也激发了他去考虑电磁的类比,导致了他在 1900 年之前就分析了矩形与圆形的波导。他是如此超前了他的时代,以至于他的工作在因为有应用而被记起时大部分已被遗忘了。就是在今天,许多从事微波工作的工程师也不知道瑞利在这个领域的贡献。

可以争论说，就需要花费的设计时间和获得的性能而言，它一定是最优的)，一个由Klopfenstein提出的设计是在给定的可容许的最差通带反射条件下情况下是最优的，因为它最短。或者说，对于给定的总长度，Klopfenstein的渐变型失配最小[15]。Klopfenstein实现这种最优是通过将失配适当地加以分布以在通带内产生等纹波的反射。这种平均误差分布(称为最小最大优化，因为最大的误差在通带内被最小化)是从将离散切比雪夫多节变换器推广到连续型极限得到的。遗憾的是，这种设计只凭手算(纸与笔)是很难完成的，尽管高效率的计算算法在1968年左右就已经存在了[16]。

这些不同形式的渐变线在很宽频率范围内可以提供对实负载(即没有虚部，译者注)的极佳的匹配。要取得对复数负载的宽带匹配就困难得多，尤其是如果要接近Bode-Fano极限。对于那些在带宽与性能都要求不高的应用，将虚部吸收入匹配网络就足够了，这样就变成了容易解决的对一个纯实部阻抗的匹配问题。然而，当负载很复杂(不仅仅是复数)时，电抗随频率的变化可能非常大而使得这种策略难以成功。我们不需要去专门设想一个具备这些特征的情形，因为仅对一个二极管混频器提供相当宽频带的匹配，就是在理论设计和实际实现上就是经常遇到的挑战。我们将以这个问题为一个例子，因为研究它以及所有可能的解决方案是很有启发性的。除了其他的考虑之外，这个问题促使我们去考虑另外几个史密斯圆图轨迹，这些轨迹能启示我们能用于大大展宽阻抗匹配的带宽的简单而有效的方法。为了理解这些宽带方法，需要扩展现有的史密斯圆图的模板库，以包括频率的变化，因为之前考虑的史密斯轨迹都是对应于在一个单一频率点上的参数的变化。

作为一个入门的例子，考虑对检波器二极管的阻抗构造一个在倍频(one-octave)带宽内的阻抗匹配说得过去的网络问题。第一个难点在于二极管是一个非线性元件，因此它的阻抗取决于其偏置点。普通情况下，可以在功耗为1mW的偏置条件下来表征混频器二极管的小信号阻抗，但是要记住在最常用的功率状况下进行阻抗表征的重要性。一般来说，这个功率并不等于本振功率。我们必须使用实际吸收的而不是入射到二极管的功率。假设一个偏置在适当水平的二极管的等效电路如图4.19所示。注意串联电感有一个小的值(并联电容的值也低)。这些在梁式引线芯片二极管中都是比较典型的值。一般来说封装好的器件其寄生参数要差得多。

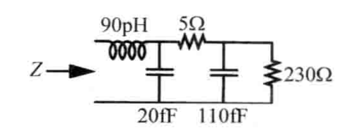


图4.19　二极管的等效线性电路模型(吸收功率为0dBm)

这个网络在6.5～13GHz频率范围内的阻抗值如图4.21的曲线 A 所示[17]。注意到在

---

[15] R. W. Klopfenstein, "A Transmission Line Taper of Improved Design"(一种性能改善了渐变型传输线设计), *Proc. IRE*, v. 44, April 1956, pp. 539-548.

[16] M. A. Grossberg, "Extremely Rapid Computation of the Klopfenstein Impedance Taper"(Klopfenstein型渐变阻抗变换器的极其快速计算), *Proc. IEEE*, v. 56, September 1968, pp. 1629-1630.

[17] 这个例子是由惠普(Hewlett-Packard)公司(现在是属于安捷伦(Agilent)公司)两个应用帖子而引起的：AN-963, "Impedance Matching Techniques for Mixers and Detectors"(用于混频器与检波器的阻抗匹配技术)与AN-976, "Broadband Microstrip Mixer Design—The Butterfly Mixer"(宽带微带线混频器设计——蝴蝶形混频器)。我们已经修改了这些技术的一些方面以得到在一个完整倍频范围内的匹配，而不只是在应用帖子中实现的1.5：1的频率比。

频带中部的阻抗已经有了所希望的电阻值，这表明通过简单地加入一个串联电感就可以使阻抗顺时针地旋转。因为等效电路里已经具备了电感，所以这种选择更有吸引力。为了计算所需要的电感值，我们仅需要找到阻抗的虚部，其实部就等于1(归一化后的值)。容性电抗的幅度等于想要达到谐振的感性电抗幅度，以去除电感的作用。对于这个例子，一个大约1.4nH的电感可以在标称的10GHz的中心频率处匹配，见图4.20。

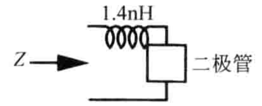


图4.20 第一轮变换：加一个串联电感

然而，旋转是随频率增加的，所以曲线 A 上的各个点随频率变化以不同的量顺时针旋转而产生曲线 B(见图4.21)。对应于最高频率 f_2 的点自然比对应于频率低端的阻抗旋转要大得多。这种行为是基于这样一个事实，即对于固定的电感值，所引入的阻抗是和频率成正比的。因为我们已经选择了在中心频率处取得近乎完美匹配的电感值，所以频率依赖的旋转导致匹配在频带边缘处恶化。这种恶化是意料中的事情因为我们所使用的是简单的单点频率匹配的方法。这里，回波损耗在频带边缘处仅是7dB，对应于SWR为2.7。这种失配的程度通常是被认为不所希望的。

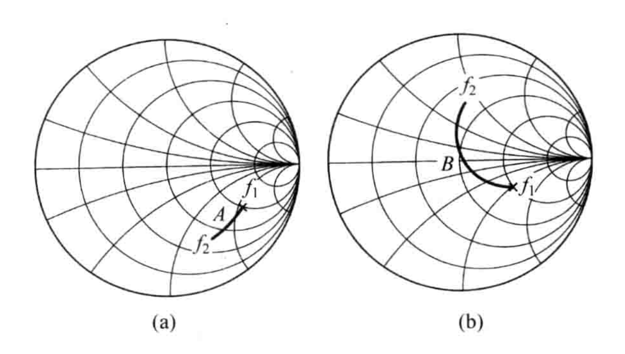


图4.21 因为增加的串联电感而引起的阻抗轨迹的旋转

让我们来考察一下提供阻抗匹配的另一个替代方法。这不是用一个串联电感，而是试一下用一并联元件。为了在标称频率下达到匹配，简单地加一个足够长的线使阻抗转到 $G=1$ 的归一化了的导纳圆上。在那里，再用一个合适的并联电纳就可以达到匹配，这就如同我们在前面的方法步骤中已看到的那样。对于这个二极管，需要加相当于43°的50Ω(特征阻抗)传输线来产生所希望的在标称频率下的旋转，而完成的单短断线匹配的第一步，见图4.22。

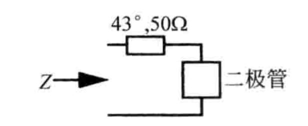


图4.22 第二轮变换的步骤1：加一传输线段

再一次，引入的线产生了频率依赖的旋转，这是因为它的电学长度是与频率成正比的。这样，如果加入足够长的线将导纳曲线旋转到在中心频率时是 $G=1$ 的圆，必然会导致高频处的过大的旋转以及在低频处的偏小的旋转。这种行为在图4.23的轨迹中可很明显看出。曲线 C 对应于加入一段传输线后的顺时针旋转结果。由于这种旋转的频率依赖性，加入一段线后使得轨迹拉长(和图4.21的曲线 A 对比)，这种行为与加入串联电感观察得到的结果相仿。

最后一步是加入并联电纳以在标称频率处实现匹配。再一次，所需要值的计算首先要

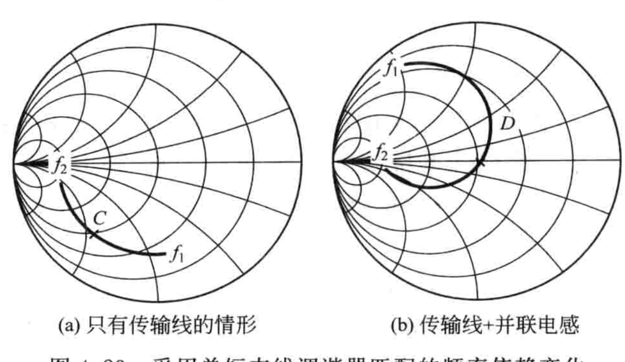


图 4.23　采用单短支线调谐器匹配的频率依赖变化

测量该曲线与 $G=1$ 的圆交点处存在的容性电纳值。然后选择电感值，使之电纳幅度等于电容的电纳。这里需要一个并联的 450pH 的电感以实现在标称中心频率处的匹配（见图 4.24）。

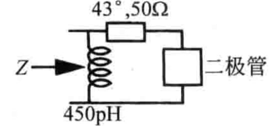


图 4.24　第二轮变换的步骤 2：加一并联电感

再一次看到，旋转是依赖于频率的，这加剧了当加入一段短传输线时已经经历到的轨迹的变长。对于加入一个并联电感的情形，在低频率下的旋转是比在高频率下的大，所以净效果是产生了标志为 D 的曲线。由于使用了两个产生频率依赖旋转的元件，我们发现使用并联方法比串联方法匹配更加恶化。事实上，最差情况下的回波损耗（在此例中主要发生在低频段处）小于 2dB，这种程度的失配在除了一些最宽松的应用情况下都是无法忍受的。这个结果强调了在选择最好的满足所有相关的优化目标之前，探索存在的几种典型方案的价值。这些考虑不仅包含匹配的质量，而且要包括可靠性、尺寸以及实现的方便性。

假如，暂且不管使用并联网络带来的令人气馁的经历，我们坚持去寻求另一种包含并联元素的方案。这种选择是我们希望的，因为需要一个 DC 回路去为二极管提供偏置。尽管有可能为串联电感的例子专门设计一个偏置电路，一个并联电感或传输线自然可以提供 DC 路径到地，这样它也可以作为偏置网络的一部分。

这种选择仍然遗留给我们一个如何在频带边缘实现匹配的挑战。采用谐振电路作为匹配元件可能会带来大的改进。定性的思想很简单：在频带两个边缘，失配可能来源于相反类型的电抗。为了实现匹配，那么需要在频带的一端加一个电感，在另一端加一个电容。单个电容或者电感都不能达到要求，只有谐振网络可以。为了更有效地利用这种思想，我们需要旋转阻抗轨迹使得在频带边缘的电抗的确是相反类型的。这个松散的描述过程仍然留下了一个问题，即如何处理电导。一个好的选择是使频带边缘处的电导等于频带中心处电导的倒数。我们将会看到这种选择将最终会导致阻抗轨迹聚集在完美匹配点为中心的附近。而且，这些电导对应的反射系数提供了对最终的最坏情况下失配的一个良好估计。

通过仔细观察图 4.23，可以看出我们需要将曲线 D 逆时针旋转一定量。因为中心频率处的电导已经匹配了，通过加一段特征阻抗等于史密斯圆图归一化阻抗的传输线以旋转曲线并不能产生所期望的倒数电导的关系。实际上，从导纳轨迹的曲率可以明显地看出，我们需要提供一个变换（实际上，重新归一化）的旋转以增加频带中心处的电导。这可以通过选择一段具有比正常归一化阻抗高的传输线来实现。选择一个特征阻抗为 120Ω，电学长度为

43°(在 10GHz 的中心频率处)的线可以产生所期望的结果(见图 4.25)。

这种将旋转和重归一化的组合如图 4.26 曲线 E 所示。注意到频带边缘处的电导实际上等于中心频率电导的倒数(一些很挑剔的工程师可能会选择更高的传输线阻抗以将轨迹再左移一些)。观察频带边缘的电纳,我们也可看到网络的导纳在低频处呈现电容性而在高频处呈现电感性。因此,选择一个合适的并联谐振器可以补偿网络在频带边缘的电纳从而实现匹配。一个过高阻抗的线可能致使末端难以被补偿,而过低阻抗的线可能会造成过补偿,致使尾部互相交叉。一个不合适的电学长度则将导致形状的不对称。

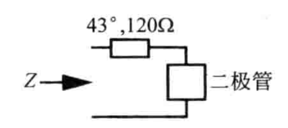


图 4.25　第三轮变换的步骤 1:加一段传输线

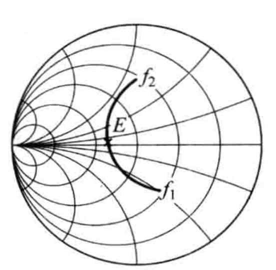


图 4.26　在加了高阻抗传输线段之后,经过旋转与电导平衡后的阻抗轨迹

注意到网络在高频边缘的感性电纳比在低频边缘的容性电纳大。这种非对称可以通过引入一个互补网络来补偿,这个互补网络的实现是通过使并联线在中心频率时具有非 90°的电学长度。这里我们在高频边缘需要提供高一些的电容,意味着是我们需要选择一个在中心频率长度大于 90°的线(这样这个线的阻抗是容性的)。我们通过并联一个电学长度 96°的 46Ω 传输线可以在中心频率处产生正确的结果(见图 4.27)。

最终的轨迹见图 4.28.我们可以明显地看出,导纳轨迹现在是分布更集中了并且大致是处于最佳匹配的周围。其结果是,失配以某种方式振荡。我们知道如果轨迹是一个完美的(译者注:圆心在原点的)圆,那么反射系数会是一个常数。其他形状必然会有纹波。如果我们能产生一个等纹波的响应,其最差的反射系数正好能符合我们对匹配的要求,我们就是得到了切比雪夫(最小化最大值)最优。尽管我们还没有十分得到那个最佳结果,很明显这里的最差情况下的失配是要比加串联电感或者单根短断线的匹配网络要小。对于这个例子,回波损耗在整个倍频范围内实际上超过 10dB,相应的 SWR 好于 2∶1。尽管这个匹配的质量并不是非常好,但在很多系统中已经是在可以使用的范围内了。

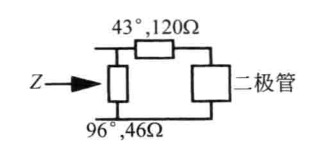


图 4.27　第三轮变换的步骤 2:加一段并联传输线

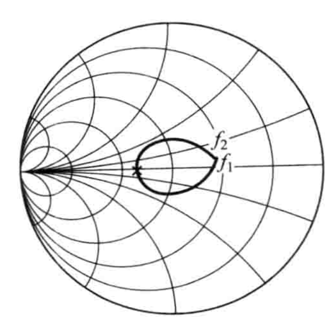


图 4.28　增加了一段短路线长度以产生聚汇之后的导纳轨迹线

在这个例子中我们研究了一系列匹配的策略，我们已经构建了一个方法（或称宏模型）集，这些方法对于在相对宽带的范围内进行匹配具有很好的通用性。尽管使用这些方法涉及一个预定假设的先验性，这个不死板的步骤仍然比那些更复杂的、形式综合的方法使用得广泛地多。这些方法并非都有正式的名字，所以我们在总结这些宏模型时，发明了一些名字。下面的大致并不穷尽所有，但是已经包含足够的操作以构成一个有用的子集了。

旋转。旋转可以通过加并联电纳或串联阻抗实现。这些导纳和阻抗可以由集总元件或者适当长度的传输线构成，就像单节短断线调谐器中的第一步那样。因为电学长度是频率相关的，所以所有操作导致的旋转角度都是依赖于频率的。

电导（电阻）均衡化。这个操作使频带边缘的电导（电阻）相等，它可能是由某些合适的旋转导致。

电导（电阻）平衡化。这个操作不仅使频带边缘的电导（电阻）相等，而且是等于中心频带的电导（电阻）的倒数。这样做可以为下一步“汇聚”做准备，正如下一段描述的。用不等于系统归一化特征阻抗的线段来实现旋转是一种同时提供平衡的快速方式。

收敛。一旦均衡化已经实现了，频带边缘两端的导纳（阻抗）可以通过加入一个符合要求的并联（串联）谐振器而达到互相靠近。因为这个谐振器在谐振频率的一边呈电感性，而在另一边呈电容性，因此它可以实现一个小小的神奇功能：将一个已经实现平衡的轨迹的两端带到一起。因为注意力是放在使实部的倒数相等，所以在频带边缘虚部可以有不同的幅度。因此为了达到轨迹端的收敛，用作终端的这个谐振器的谐振频率不一定就是中心频率，而可能是高于或低于中心频率。通常收敛是倒数第二步，而不是最后一步。因为线的阻抗实现起来可能受到限制，所以基本不可能只用一次操作就在史密斯圆图的中心实现一个收敛的轨迹。最后的一个旋转或变换（这再一次是通过加一段适当长度的有正确特征阻抗的传输线来实现的）可以完成了匹配。

最后，注意到这里叙述的一般步骤并没有直接给出在通用形式下所能达到的失配指标。因此，不能保证一定成功。尽管这么说，执行这个设计步骤（以及加以实际实现）是十分简单的，人们经常会先试试这个方法（译者注：再去考虑其他的办法）。幸运的是，用这种方法一般都能得到可以接受的结果。

第5章 连接件 电缆线与波导

5.1 引言

虽然本书所关注的重点是实现分立的平面射频电路，我们在此考虑了几种重要的、本质上是三维的元件：连接件、电缆线与波导。我们将看到这些元件的有用频率范围部分是被模(mode)所限制的，而这些模是物理尺度的函数。另外，还要考察这些用不同方式将射频能量从一个地方传至另一个地方的衰减特性。

5.2 连接件

5.2.1 模与衰减

为了对尽量大带宽的信号有最平坦的响应，一个 RF 连接件应该在它的整个长度上表现出相同的阻抗。这一点可以通过统一连接件各处的尺寸，并在其中填充同样的电介质来保证。虽然这点要求看起来显而易见，但是我们可以看到至少有一种极为流行的连接件不能满足这个要求。

最好的同时是最通用的射频连接件采用的都是同轴结构。同轴结构最重要的特点就是它们拥有自屏蔽特性，这样辐射损失就不再是一个问题。但是，有一点必须要特别注意，即要维持横向电磁(TEM)波的传播。你或许可以从大学电磁学的课本中回忆起，对于一个 TEM 波，$\boldsymbol{E}$ 和 $\boldsymbol{H}$ 都没有与传播方向平行的分量。从而在足够高的频率下，非 TEM 波的传播也会发生。同时，存储在高阶模的能量或是以高阶模传播的能量会引起阻抗的急剧变化。随着频率的上升，这种高阶模式被激发出的趋势也越强。在实际的传输发生前，在这些模式中储存的能量自然表现为不作功的能量，从而会影响阻抗。实际的传播则造成了电阻型的扰动。

非常粗略地讲(这需要对 Maxwell 教授表示歉意了)，当波长到足够小了以至于波能通过结构的某个关键尺度而“挤”过去的时候，波导高阶模式传输就发生了。从而，如果想防止这类的传播，那些导体的尺度就必须足够小，以抑制除了 TEM 以外的其他模式的传递[①]。对于我们正在考虑的同轴电缆，有关的关键尺寸大约就是内外两层导体之间的环形区域的平均电周长。一个非 TEM 波(特别是，TE_{11} 模式)，在其波长等于这个平均周长的时候，它就可以传播了[②]。第一个这样的高阶模就具有下面给出的波长的近似值：

---

① 在专业的语言里，模(mode)这个词可以作为动词用。*To mode*(模)意味着产生高阶的模。这样，我们就要担心在该频率下一个连接件或传输线开始“to mode”的频率。是的，这看起来有点滑稽(用名词作动词)，但是 *c'est la mode* (法语，意指是一个风尚)

② 如果有关模的专用术语对你来讲不熟悉，不必担心。5.4 节会详细地讨论这个议题。

$$\lambda_c = 2\pi\left[\frac{a+b}{2}\right]\sqrt{\varepsilon_r} = \pi(a+b)\sqrt{\varepsilon_r} \tag{1}$$

其中，a，b 分别是导体壁之间空间的内径和外径，ε_r 是空间内填充物质的相对介电常数；如图 5.1 所示。b/a 的值处于 1～15 之间时，该近似的误差在 3%以下。

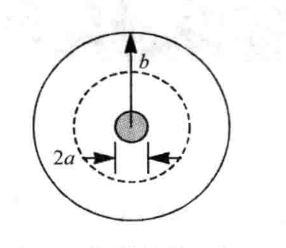


图 5.1 同轴结构的截面

人们通常希望能进行没有高阶模的运作，这就解释了为何随着工作频率范围的提高，连接件和电缆逐渐地变得越来越小。一些常见的同轴连接件有：UHF，N，BNC，TNC，SMA，SMB 和 SMC，如图 5.2 所示。



图 5.2 各种同轴连接件(所示大小与实际尺寸并不成比例)

这些连接件中最早的是 UHF，它早在 20 世纪 30 年代就由 Amphenol 公司的 E. Clark Quackenbush 为无线电广播事业开发出来了[③]。它通常应用于业余无线电。其插口型号是 SO-239，对应的是 PL-259 插头。遗憾的是，尺寸庞大的接口实际并不适合现在定义的 UHF 的工作频率范围[④]。事实上，它有一个严重的设计瑕疵(沿接口方向阻抗不等)，从而导致它的特性到了 VHF 的中高端范围(30～300MHz)明显的下降。50MHz 时的损失已经超过了 1dB[⑤]。从而，这种接口不应该用于高于 100MHz 的情况。坦白地讲，它应该被淘汰掉了。但是，它依然应用于很多 2m(指波长，译者注)的 FM 业余无线电设备，甚至偶然也应用在了一些 70cm(440MHz)的无线电设备上，尽管其代价是性能大为下降。它唯一的优点就是便宜。我们邀请大家一起来让这种连接件消亡吧！

N 接口以它的设计者、贝尔实验室的研究员 Paul Neill 来命名。自从第二次世界大战以来，它被广泛应用于 RF 测试设备。它的尺度使得它的无高阶模的频率上限提升到了大约 18GHz[⑥]。但是，最常用的产品在外层导体内有一些沟槽，这种型号通常具有的指标为 11～12GHz 作为其上限频率。没有沟槽的产品性能更好些，直到 18GHz 性能都很好。请注意不是所有使用 N 接口的电缆都能在整个频率范围良好工作。

另一个问题是，根据设计，在接口两头中心针脚与插座的连接处在两个连接导体的肩部

③ 当时，Amphenol 还叫 Phenolic。

④ UHF 所指的频率范围随着时间的推移发生了变化，但这个连接件的名字却没有变。

⑤ 再一次提醒，“仅是”1dB 的损耗就会是 21%那么大。

⑥ 一些参考资料则认为 N 代表“Navy(海军)”。

有一个小的间隙(大致 75～100μm)，如图 5.3 所示。外层(接地端)导体(没有在图中画出)被设计成首先对接，这样就留出了给中心导体的空隙。做这种设计选择的原因是要求过紧的机械加工容差与低的膨胀温度系数以消除这个缓冲空间的做法是不实际的。如果中心针脚连接时不保留这样一个缓冲空间，将连接件阴阳两件插紧到外部表面对上几乎肯定会损坏中心针脚。与尺度实际变化相关联的小的不连续性引起了频率增加时阻抗的变坏。正如可以从图中所猜测的那样，导体直径的减小主要是与一个寄生串联电感相关⑦。

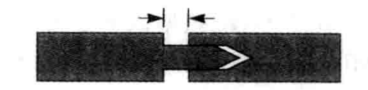


图 5.3　当外部的导体对接上后的中心管脚的详图

就像所有其他螺口接口一样，千万不要使用钳子来收紧一个 N 接口，那样的话中心插头迟早会损坏(同时接口也将刮伤)。想达到最大寿命，推荐用规格为 12in・lb(135N・cm)扭矩的扳手来拧紧。除此之外，在使用次数不多的前提下，用手拧紧也是可以的。其他需要注意的事项包括：绝对不能将润滑油(或其他基于石油的溶剂)涂在接口或电缆上。如果接口很难正常拧紧，那是因为被错扣了，就要将它们换掉。涂润滑油不但不能修复这个坏的部件，还会进一步损坏接口和线缆的塑料部分。如果连接件需要清洗，那么要使用高压空气或者专门设计用来清洗接口的溶剂。

BNC 接口，以发明人 Neill 和 Carl Concelman(Amphenol 公司)的名字来命名。由于它的方便性和相对较小的尺寸，它被广泛应用于 RF 装置中。为了能够快速连接或切断连接，插刀式的(bayonet——刺刀，BNC 中的“B”⑧)接地连接方式替代了螺纹连接方式。它们的内部尺寸是能够保证至少到 18GHz 的无高阶模式操作，但是卡口严重的不稳定性在工作频率大于 1GHz 时就开始影响性能了。一般而言，上截止频率限取为大约 3～4GHz，虽然 BNC 接口经常用于高于这一频率的情况——但代价是不断变差(及不可预测)的性能。震动会调制阻抗，导致其他的负面效果，这在移动应用领域中更加明显。这个问题通过采用 BNC 的螺纹接口得到了解决。TNC(threaded Neill-Concelaman，于 20 世纪 50 年代后期开发出来)适用于至少 11～12GHz 的情况。最好的情况可以一直用到 18GHz。BNC 接口也像 N 接口一样，同时 SMA/B/C 接口(下文将提到)也是如此，在其内部针脚的肩部与其插座间存在着一个小间隙，导致电抗不连续。这个不连续性随工作频率的上升变得更严重。

20 世纪 60 年代，对更小尺寸接口的需求促成了 SMA(sub-miniature, type A，超小型 A)接口的出现。它比 BNC 尺寸要小得多，当用于与半刚性的同轴电缆(通常工作在 12GHz 下)连接，可以在至少大约 18GHz 下工作。如果加以小心，还可以工作到超过 25GHz。用规格为 56N・cm(5in・lb)扭矩的扳手拧，可以保证接口的性能长期不至于变差。用手来拧则很难达到这一扭矩，而用普通的扳手则很容易拧得过紧。

另一种小型螺纹接口是 SMC 接口。它比 SMA 要便宜，同时也要小一些。SMC 接口

---

⑦ 更加复杂的模型也考虑到电容的边缘效应，其对这个不连续性的建模是一个 *CLC* π-模型。无论是哪一个模型，不连续性降低了工作频率的范围。

⑧ 这是一个通常的说法，而且这个说法为许多教科书的作者采用，“BNC”代表“baby(婴儿)N connector(连接器)”。但是 BNC 的确代表“bayonet Neill-Concelman”，以与带螺旋的 Neill-Concelman(TNC，而不是依然有些人认为的“tiny N connector”)区分开来。这个错误的概念源自 *Technique of Microwave Measurements*(微波测量技术)(MIT Rad. Lab. Ser., vol 11)，p. 10，尽管这本书通常是个可靠的来源。注意，还有一个不太被知道的 C 连接件。这个连接件是由(你大概可以猜到了)Concelman 开发出来的，并以他的名字命名。这个 C 连接件看起来就像 BNC，只不过是大一些。

可以在10GHz频率以下正常工作。适合的扭矩在3～4in·lb(34～45N·cm)范围之内。SMB接口是SMC接口的卡口版,它的工作范围下降到了大致4GHz(与BNC差不多,也有着类似的原因)。

接口的插入损耗取决于它的连接方式,是直连还是直角连接;采用的是螺纹还是卡口连接。但是,由于电缆上的损失几乎总是占据主要地位,所以衰减的精确值一般是不重要的。不过,我们可以给出粗略的估计。根据一个非常粗略的经验法则,可以假设,卡口直连方式和螺纹直角连接方式的接口,相较于其他都相同的螺纹直连方式接口而言有两倍大的损耗(以dB为单位)。对于较大的接口,每单位长度的导体损耗较低,但是总长比较长,从而每个接口的插入损耗区别不大。随着频率的增加,以dB为单位的损耗通常会介于正比例增长和平方根增长之间。在工作频率为低GHz范围,N接口和SMA接口的损失典型值通常在大约0.04dB/$\sqrt{\text{GHz}}$以下(对直连形式),大约0.06dB/$\sqrt{\text{GHz}}$以下(对直角连接形式)。生产厂商在产品手册中给出的指标一般是相当保守(即较差)的数值(一方面是为了保险,另一方面低指标可以让厂家不必对产品进行测试)。BNC和TNC一般对直连和直角连接方式分别有低于0.1dB/$\sqrt{\text{GHz}}$和0.15dB/$\sqrt{\text{GHz}}$的损耗。由于不那么可靠的接地方式,BNC的损耗会比TNC的略大一点,同时随频率的增长也会更快一点。对于SMC接口,相应的值是相似的,大约为0.15dB/$\sqrt{\text{GHz}}$和0.2dB/$\sqrt{\text{GHz}}$。SMB的损失大约为SMC的两倍(以dB为单位)。由于每个厂商给出的数值有很大的波动,这里给出的数据仅仅是一个粗略的近似。

对于必须承受经常性连接断开的设备,必须要有一种不同的接口。普通的接口受频繁插拔的影响,会使得中心插头性能下降。正如前文所述的,生产过程考虑要求在中心插头的匹配结构上增加一个小的间隙,但致使随频率上升造成了不匹配问题更加严重。为了避免这些问题,Amphenol和惠普公司改进了一种德国设计,制造出了一种无极性接口,它的中心和外层导管采用一种“头顶头”的接触形式(与其他接口的“针脚对插座”的形式相反)。APC-7(7mm “Amphenol precision connector” 或者 “a precision connector”——一种精密的连接件)接口广泛应用于工作频率至少是18GHz的设备中。

为了连接两个APC接口,首先,顺时针旋转第一个接口的外层套管直到螺纹出现。这个套管被保持松开状态。接下来,将两个接口对准,并把另一个接口的套管在刚刚松出来的螺纹上收紧。在收紧完毕时,两个外层套管不应该接触到一起;否则,它们会使中心插头的匹配接触面分离。最重要的一点是在收紧过程中,只能旋转外层套管。如果匹配接触面之间产生了互相摩擦,那么产生的磨损就会损坏接口。当然,也不允许直接接触匹配面。显然,这类接口的性能取决于接触面的清洁度以及机械误差的精确度。为了减小损伤并得到较为平滑的随频率变化的阻抗变化曲线,合适的扭矩(135N·cm,或者12in·lb)是必须的。在没有力矩扳手的情况下,用手收紧也是可以的。

对于在更高工作频率下的操作,我们需要小的尺寸来避免高阶模式的发生。3.5mm,2.92mm(被称为Wiltron K接口,被用于可覆盖的最高频带),2.4mm,1.85mm(或者V接口,同样是因为所覆盖的频段)和1.0mm(W)接口的工作频率分别为大约33GHz,40GHz,50GHz,65GHz和110GHz。除了1.0mm接口,所有的这些接口都应该被收紧到90N·cm(8in·lb),1.0mm接口的扭矩则规定为34N·cm(3in·lb)。再强调一次,使用力矩扳手是保持接口良

好特性的“绝对”强制条件，这些特性包括典型的损失在 0.1dB/$\sqrt{GHz}$之内。工作频率越高，对力矩扳手的要求也就越高。

关于这些接口的最后一个说明是，由于直径和螺纹间距通常是兼容的，SMA 与精确型(APC)3.5mm 接口及 2.92mm 接口可以互相连通。但是，SMA 采用的是特氟龙介质，而精确型接口采用的是空气介质。作为补偿，SMA 的外层套管要薄一些。另外，某些供货商提供的 SMA 接口没有很高的制造精度。这些方面联合导致了在不同接口互相连通时，即使仅仅连接一次，也将造成很高的损坏率的情况。当更换接口是一件很费劲的事时，作为一个额外的一种保护措施，可以在 3.5mm 接口加接一个短的延长器(比如说，公口在一端而母口在另外一端)作为一个牺牲的接口。这样，牺牲接口将承担所有的损伤，这使得损伤的清除或替换变得容易。很多厂商都会提供这种“接口保护器”作为牺牲接口。

近年来，紧凑型消费无线设备——蜂窝电话、GPS 定位仪、无线 LAN 网卡等的(以及正在进行中的)增长，带动了全新一类的超小型与超低成本接口的研发。幸运的是，无线传播的物理机制所要求的低于大约 5～10GHz 的工作频率在某种程度上降低了设计的难度。有代表性的应用在这些设备上的接口有 MCX 和 MMCX，都是卡口类型(与 SMB 十分相似，但是要小得多)，额定工作上限为 6GHz 左右。对于直连和直角连接的插入损失通常是分别好于 0.1dB/$\sqrt{GHz}$和 0.2dB/$\sqrt{GHz}$。

很多无线 LAN 连接点上十分流行采用 TNC 和 SMA 接口的反接版本。选择这种连接可防止与天线的不小心的误连接。这种连接可能会导致辐射能量超过规定的上限。不可避免的，这样的结果鼓励了生产合适接收器的小规模的产业。

75Ω 连接件

上文中关于接口的描述都是关注于标准的 50Ω 接口的，还有少许一些标准则是针对 75Ω 情况的。例如，F 接口可以说是 75Ω 应用领域(比如有线电视)中使用最广泛的接口了，而且没有相应 50Ω 这样的接口。遗憾的是，有关中心针脚的尺寸规定是如此之宽松，以至于近似于荒唐地容易来生产一种性能很差的接头，或是直接造成损坏：中心插头允许的直径范围可以从 0.5mm 直到超过 1.6mm！更糟的是，有时还能够发现(由不可信任的厂家生产的)电缆的中心导体直径甚至超出这个已经是十分宽松的范围。不用说，试图去连接两个处于允许频率范围两个极端的接口，毫无疑问会造成接口的损伤或根本就连不上。大部分 F 接口的工作频率上限是 1GHz。

几种其他的 75Ω 接口是根据 50Ω 标准的原型设计出来的，所以我们可以在 75Ω 环境中找到 N，BNC 接口……，虽然它们比 50Ω 环境中的对应接口要少见得多。很重要的一点是中心针脚的尺寸在 75Ω 和 50Ω 阻抗标准下是不同的，所以尝试着连接公的 50Ω 和母的 75Ω 接口将会毫无疑问地损坏至少其中的一端。

5.2.2 非线性效应

因为在连接件中仅涉及导体和线性度很好的电介质(例如空气和特氟龙)，提到接口会表现出非线性效应看起来很奇怪。事实上是，的确有可能存在非线性效应，并造成了明显的失真(甚至是谐波和交调乘积项的寄生辐射)。因此，如果工程师不知道这种可能性，当有问题发生时，会花去很多时间来找出问题的原因所在。

造成接口的非线性一般有两种原因。第一种归咎于普通的腐蚀，很多金属氧化物、氯化

物和硫化物都是半导体，腐蚀可以造成一种或者多种这类的化合物（即半导体，译者注）。非线性可以与这些化合物产生的肖特基势垒或隧道势垒联系起来。一个明显的补救措施是将所有的接触面都保护起来，以防腐蚀。但是，这一简单的处理方法实际上很难实现，尤其是对于室外的设备。使问题更为复杂的是：导体表面经常会镀一层银来降低高频损失，而银却是很容易被钝化的[9]。

另一个产生失真的原因是在大信号条件下可能饱和的磁性材料。磁性材料通常不会被有意识地使用，但是镍和不锈钢（经常被用于抵抗腐蚀的镀层）在这里通常是肇事者。很多接口可以有不同的涂层，这样的接口应该用在连低的交调失真都不容许的场合。

5.3 同轴电缆

5.3.1 为何要同轴

上文对同轴电缆接口的讨论没有回答一个基本的问题：为什么要用同轴线？答案或许很明显，但还是要说明一下。在低频条件下，任何旧款的线缆都可以（而且一直是）在损失很小的情况下，将信号从一个地方传到另一个地方。例如，双绞线被广泛应用于数据通信（比如，现在广泛使用的“CAT-5”电脑网络线缆）。双绞保证了两条导线受到的由杂散场和其他的导体的邻近效应产生的扰动的影响相同，这样就保证了平衡的特性。辐射损耗同时也得到了抑制，允许双绞线能在 100Mbps 和 1Gbps 的数据速率下正常工作。目前正在进行的工作就在尝试将数据速率再提高一个数量级。

确保扰动处在共模模式毫无疑问是有益的，但是当频率升高时，这种无屏蔽的结构开始表现出天线的行为。我们在有关天线那一章会介绍，当导体的长度与波长处于同一量级时（即十分之一波长），辐射将变得很显著。这种能量的辐射导致了只有部分能量到达了预期的目的地，即衰减增加了。

而具有屏蔽层的同轴结构就能避免辐射损失。避免产生高阶模式总是一个被关注的问题，并且我们已经看到了用这种抑制高阶模的希望来解释为什么在使用频率上升的情况下，接口和电缆的尺寸不断地在减小。至于说到衰减，缺乏辐射损耗就意味着只需要考虑导体和电介质的损耗。这两种损耗机制可以从对频率不同依赖关系进行区分。趋肤效应导致的导体损耗随频率的平方根而变化，而电介质损耗通常随频率线性变化（或更快）。

为了证实这些论断，并更细致地研究这些机制，我们来看一下下面这个衰减因子的表达式[10]：

$$\alpha \approx \frac{1}{2}\left(\frac{R}{Z_0}+\frac{G}{Y_0}\right) \tag{2}$$

其中，α 是每单位长度以奈培（neper）作为单位的衰减值[11]，R 是每单位长度的串联电阻，

---

⑨ 与普遍看法相反，银的钝化主要是由硫化银导致的，而不是氧化银。同时，铜上的铜绿通常是氯化物（近海地区）或者硫酸盐（内陆地区），而不是氧化物。

⑩ 见第 2 章，公式（37）

⑪ 可以回顾一下，幅度是每增加一个 $1/\alpha$ 的距离衰减 e（自然对数的基）倍。奈培与分贝因此是互相成正比的，一个奈培等于 20lg(e) 或大约 8.69dB

G 是每单位长度的分流电导，Z_0 是特征阻抗，Y_0 等于 $1/Z_0$。

公式(2)右边第一个项是导体损耗。如果我们假设工作的频率都能保证趋肤深度远远小于导体厚度，就可以假设电流均匀地流过一个厚度为表面深度 δ 的薄的圆筒来估计 R 的值

$$R \approx \frac{1}{2\pi\delta\sigma}\left[\frac{1}{a}+\frac{1}{b}\right] \tag{3}$$

其中，σ 是导线的电导率，δ 是趋肤深度

$$\delta = \sqrt{\frac{2}{\omega\mu\sigma}} \tag{4}$$

从而，

$$\alpha_c = \frac{\left[\frac{1}{a}+\frac{1}{b}\right]\sqrt{\frac{\mu}{32\sigma}}}{\pi Z_0}\sqrt{\omega} \tag{5}$$

可以看到，正如上文所断定的，导体产生的衰减 α_c 随频率的平方根而变化。同时我们还可以知道，自然地，导体直径越大，或电导率越大，衰减系数就越小。

对电介质损耗的频率依赖关系可以通过想象损耗产生的原因是一种偶极子的“摩擦”来定性地预测。如果我们假设必须花费一定的能量去克服这种摩擦而使得偶极子极性逆转，这样能量损失的速率(因此也是功率损耗)就应该与频率成正比。因此预期由这种机制导致的衰减正比于频率。作为一个留给读者的练习，可以证明因为电介质损耗造成的衰减常数 α_d 的表达式为

$$\alpha_d = \frac{[\tan\delta]\sqrt{\mu_r\varepsilon_r\mu_0\varepsilon_0}}{2}\omega \tag{6}$$

其中，$\tan\delta$ 是电介质材料的损耗角正切(有点不巧的是，趋肤深度和损耗正切因子的表达式正好都使用了符号 δ)。不管怎么说，我们看到，在以奈培与分贝为单位的情况下，电介质引起的损耗确实随频率的增长而线性增长[12]。注意，与导体引起的损耗不同，电介质损耗与导体尺寸无关。

通过上述的推导可以发现，在低频的情况下，对实际采用的电缆，其衰减因子大致在低频率下是个常数，而后，随频率的逐渐上升，当导体损耗引起的衰减占据主导地位时衰减因子随频率的平方根增大。随着频率的进一步上升，衰减因子转变为与频率成正比地增大，这时电介质损耗占据了主导地位。有些电介质材料的损耗角正切随频率上升而增大，因此衰减因子的增长快于频率的一次幂。电缆的截止频率上限通常选在电介质损失开始占据主导地位的点附近。

通过以上的分析，我们可以看到，以空气作为电介质，采用最大物理尺寸的电缆，可以提供最少的衰减。但是，由于要考虑高阶模产生的可能，电缆尺寸存在着上限。作为一个结果，对于任意给定的频率，所能达到的衰减因子都存在着一个下限。

5.3.2 同轴电缆的种类

对于低频情况下的一般应用(比如说，低于 1GHz 左右)，通用系列的普通的软同轴电

[12] 大多数用在同轴电缆上的电介质材料都很好地满足这个假设。

缆，标识为 RG-n/U。其中字母“RG”代表“无线电波导”(radio guide)，“U”代表“通用”(universal)⑬。这个型号标识最初包括了硬线和波导，但是近来的应用都主要在低频同轴电缆上。这一系列电缆包含了不同的尺寸和不同的电介质，见图 5.4。许多公司还提供其他改进型的电缆，包括通过使用更好的电介质(例如 PTFE 或聚乙烯气泡垫)以及更优的导体和外层材料。5.6 节给出一个有代表性的例子的特性。

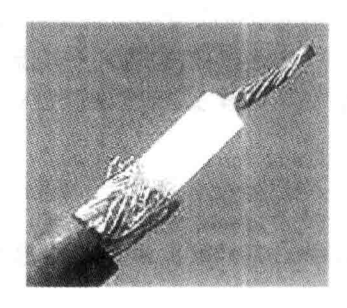


图 5.4　典型的同轴电缆

普通的软同轴电缆有一个问题就是：反复弯曲电缆会造成导体和绝缘体的疲劳，致使性能下降。另外，软电缆很容易被弯曲成陡的直角，以至于形状变形，从而导致电学特性下降，如果电缆被打结的话也许会造成永久破坏。半刚性和刚性的电缆就是为解决这些问题而研制的。半刚性同轴电缆通过采用有些柔韧性的外层导体保持了一定的柔软性，同时允许电缆保留成弯曲成形的样子。可以采用一些特殊的夹具来避免扭结。就像通常的软性同轴电缆(或其他任何线，就这一点而言)一样，过紧的弯曲半径会降低电气性能。刚性同轴电缆(也被称为硬线)，正如其名字说明的那样，完全不能弯曲，从而厂商提供了预先做好的各种形状以及各种长度。

半刚性电缆通常使用特氟龙材料作为电介质并通常有两种型号：0.141 英寸直径和 0.085 英寸直径。同时也有 0.141 英寸特氟龙软电缆，它的特性介于半刚性和大直径软电缆之间。

75Ω 电缆

因为大多数射频器件和系统都是 50Ω 阻抗的，我们关注的重点也就是 50Ω 接口和电缆。但是介绍一些关于 75Ω 的连接件也是需要的。

正如已在其他地方解释过了的⑭，功率处理能力与衰减的折中选择使得 50Ω 成为了标准阻抗。对于以空气为介质的同轴电缆，阻抗在 30Ω 时达到最大功率处理能力，77Ω 时则能产生最小衰减。这两个(经过取整)值的平均被选作一般情况下的折中结果。

但是在某些特殊的应用中，比如对于有线电视，为减少必须配置的但价格昂贵的中继器的数目，达到最小的衰减是非常重要的。在这种应用中功耗是比较低的，所以功率处理能力并不重要。这样，75Ω 阻抗就成为了视频产品的标准。

几种常见的 75Ω 电缆特性也将在表 5.1 中列出(5.6 节)。

5.4　波导

我们已经看到，同轴电缆的衰减在低频下由导体损耗主导，在高频下则由电介质损耗所影响。我们已经知道减小衰减的办法是采用空气作为电介质，并不受限制地增大电缆的尺寸。但是，考虑到高阶模被激发的可能，我们不能这么去做。而使用波导的目的，就是为了回避这一局限。

我们可以假设，与其通过采用越来越小的几何尺寸去尝试阻止能量的高阶模式传播，还

⑬　G. L. Ragan, *Microwave Transmission Circuits*(微波传输电路)(MIT Rad. Lab. Ser., vol. 9), McGraw-Hill, New York, 1948, p. 244.

⑭　见 3.5 节。

不如去实际地利用它。从而，我们选择不用TEM波进行传播。基于这种选择，可以自由地选择更大的导体尺寸。进而，因为高阶模式不含直流分量，所以同轴电缆中的中心导体也可以去除，这样就得到了称为波导的结构。通过刚才的描述，很明显波导就是一个空心导管。最初，使用圆形空心管实现波导传播(在20世纪30年代，由贝尔实验室的工程师George C. Southworth和MIT的Wilmer L. Barrow独立而近乎同时完成)，但是也有截面为椭圆、长方形、正方形的波导也被使用[15]。长方形截面波导是最容易计算的(在所有结构中)，从而，长方形波导成为了使用最为广泛的波导，同时也是在我们的这个简略的讨论中唯一涉及的波导结构。若想详细了解其他截面的波导模式，有兴趣的读者可以参考Ramo，Whinnery和Van Duzer合著的一本很好的教材《现代通信系统中的波和场》(Wave and Field in Modern Communications Systems)(第三版，Wiley出版社，New York，1994)或者E. C. Jordan著作的《电磁波和辐射系统》(Electromagnetic Waves and Radiating Systems)(Prentice-Hall, Englewood Cliff, NJ, 1950)。

5.4.1　有关模的术语

我们已经使用过了TEM这一术语，它代表了电场和磁场方向都垂直于传播方向的一种模式。单根导体的波导不能传播TEM波，因为在TEM波中含有直流分量。但是这个结构可传播TE波或TM波，至于是TE还是TM波则取决于波导是如何被激励的。在标准的术语体系中，“T”后面跟的字母描述了垂直于传播方向的场的种类。完整的模标记还需要标出此种模在长方形波导中x，y方向上截止模式的半波数，对于圆形波导则对应着圆周方向和直径方向的半波数。这样，$TE_{m,n}$告诉我们在一个长方形波导内，在x方向的电场有m个半波，而在y方向的电场有n个半波。

考虑一个具体的例子，长方形波导中TE_{10}波为截止模的情形。这种模式有一个横向(即垂直于传播方向，译者注)的电场，其在左右壁上的值为零，在中心处的值最大(即在x方向上有一个半波)。同时这个电场在垂直方向上(y方向)没有变化。

下面看另一个例子，圆形波导的TE_{01}波为截止模：电场在沿圆周方向上没有变化，而且在波导的表面，其值为零。同时只有一个峰值，就是在波导的中心。

虽然波导可以支持无数多种模式，但是实际情况下，通常会限制工作频率的范围，以确保只有一种模式的波在传播。这一要求与同轴电缆中应用的基本理念是相同的：当从较低的频率上升到一种模式的截止频率时，无功电能储存上升。一旦一种模式真正开始传播，则会增加一个实数项。这两个效应都会造成阻抗改变(或者，按照更加适合波导领域的语言来说，SWR变差)。通常都会使用最低频率的模式，所以波导的尺寸必须能够满足操作频率范围的要求。像同轴电缆一样，随工作频率的上升，波导的尺寸也要下降。但是由于波导拥有较大的导体表面积，所以其在衰减方面的性能还是有了很显著的改进。

在高的频率下，通过将波导尺寸缩小到足够小来抑制高阶模式是不现实的。在这种情

[15] 第一个发表波导传播理论的是瑞利爵士，这是在1897年。然而，他相对于他的时代是过于超前了。在一个工程师们还不能产生能在尺寸不大的管道中传播信号的时代，没有人会在乎波导的传播。几乎是40年之后，Barrow与Southworth在相互独立的情况下(并且他们也不知道瑞利的先前工作)进行波导的工作。直到他们发表各自的结果时，他们才知道对方也在从事同样的工作。

况下，使用高阶模式进行传播是更有意义的。这种过模(overmoding)在毫米波领域就很有意义。过模滤波器、波导和其他电路在近 THz 电子学领域都是活跃的研究课题。

5.4.2 波导的衰减性质

对于矩形波导中的 TE 和 TM 两个模式波而言，因为趋肤效应的原因，损耗最终随频率的平方根而增加。这个原因是与 TEM 波在同轴电缆中的传播情形一样的：频率的增加引起了被感应出来的表面电流能够流通的有效体积减小了。这种损耗机制看起来是如此的根本，以至于当你听到这种损耗是可以被避免的(至少是理论上可以)时候会感到吃惊。对于随频率上升而造成的趋肤深度下降我们无能为力，但是如果相关的表面电流能够随频率上升而下降的话，那么净损耗就真的可以下降。

令人惊讶的是，这就是的的确确 $TE_{0,m}$ 波在圆形波导中发生的情况。这些模式波(它们在矩形波导内不能传播)与其他波不同的特点是没有垂直于导体表面的电场分量。理论上，在频率趋近于无穷的情况下，这种模式波的衰减趋近于零。现实中，弯曲、表面的粗糙、与理想轴对称的偏离，都会产生其他带有损耗的模式波。但是，如果格外小心，相较于其他模式波而言，我们可以在一个有用的、很宽频率范围内取得非常小的衰减。在发明光纤之前，使用圆形波导是传递信号损耗最低的方式。

理论的推导当然是很好的，但是来说明波导确实能具有特别低的衰减特性的最好方法是去看一下一些实际的数据。直接可以买到的工作频率在 10GHz 下的波导完全能达到每 30m 0.5dB 衰减的性能。而最好的软电缆只能做到 5GHz 下 13dB 的衰减。也可以在 100GHz 下得到 15dB 衰减(同样的是，每 30m)的波导。作为比较的最后一个例子是，光纤在 400THz 下的衰减小于 1dB/km。

5.5 总结

我们已经看到，如果想要得到恰当的工作状态并长时间内维持这种状态，必须很小心地选择和处理连接器。对于高阶模的担心和机械尺寸的均匀性都会影响到工作频率的范围。这就解释了，比如说，尽管 BNC 和 TNC 两种接口有同样的关键几何尺寸，但却有不同的频率上限。

进一步考虑到产生衰减的原因，这些同样的说法也可应用于同轴电缆。可以看到大的尺寸可以使导体损耗不成为一种主要的损耗机理，但是因为无高阶模式工作的要求，采用增大尺寸这个方法有其最终的限制。

通过采用那些不需要要求尺寸远远小于波长的结构可以来缓解这些限制。事实上，对正常的工作，我们看到为了保证高阶模式波的传播，波导的尺寸必须大于一个半波长，它们从本质上讲是个高通结构。这个性质解释了为什么 AM 无线电信号(波长范围 200～600m)基本上进不了公路隧道(highway tunnel)，但是 FM 信号(它的波长范围在 3m 以内)就可以相当不错地在里面传播。

具有较大的表面积和能够消除电介质损耗这两方面的原因使得波导表现出极低的损耗。较大的波导尺寸通常也能承受更高的击穿电压，因此波导的功率处理能力也要优于同轴电缆。

5.6 附录 同轴电缆的性质

表 5.1 列出了几种相当流行的同轴电缆，以及它们的特征阻抗与在四种不同频率下、长度为 30m 的电缆典型衰减值。注意，对于所有列出的同轴电缆，衰减值在低频时确实与频率的平方根成正比（毕竟，理论的预测还是会对的）。表层损耗直到 5GHz 都是主导因素，而电介质损失则可以从损耗曲线斜率的不太显著的增加中看出来是起了作用。

表 5.1 一些柔性同轴电缆的近似衰减值及特征阻抗 Z_0

类 型	Z_0	（在下列频率下）每 30m 的损耗/dB			
		100MHz	400MHz	1GHz	5GHz
RG8/U	52	2	4	9	30
RG58/U	52	5	11	18	60
RG174/U	50	9	18	31	100
RG213/U	50	2	4	9	30
9913	50	1.4	2.7	4.5	13
RG6/U	75	2	4	6.5	—
RG11/U	75	1.3	2.6	4.3	—
RG59/U	75	2.5	5.1	8.2	—

不同种电缆之间的衰减特性区别很容易解释。RG174 的外径是表中最小的，仅有 0.1 英寸（2.5mm），因此其衰减值是最大的，这是因为导体损失起主要作用。由于其较大的衰减值，RG174 的使用被限制在相当短的距离。

在另一个极端是 9913，其极低的损耗归结于它的大尺寸。中心导线是相当粗的 9.5 口径（gauge）（0.405 英寸直径，大约 1cm）导体（能与 UHF 接口匹配的可以选择的最大尺寸）。由于它的笨重体积，9913 既不容易也不应该进行弯曲。这种电缆在所有通常可以买到的软电缆中损耗是最小的，但它也是最早会开始产生高阶模的。

体积稍小一点的是 RG8，被广泛应用于业余无线电领域。中等的尺寸决定了它中等的损耗。需要注意的是，其泡沫电介质的物理强度不好。反复的弯折电缆会造成中心导体移位，从而产生阻抗不连续的现象。极端的移位甚至可以造成短路。

对于一般目的的应用，RG58 或许是使用最广泛、频率范围直到 VHF 的同轴电缆，因为它比 RG8 重量轻、损耗较低，并且有合适的功率处理能力。注意 RG58（“细电缆网络”）在以太网实现早期取代了 RG8，而它自己后来又被双绞线所替代。

半刚性和刚性同轴电缆用于需要最大的机械稳定性和低损耗的情形中。表 5.2 列出了几种半刚性同轴电缆和硬线的代表性数据（依据外径排序）。0.141 英寸特氟龙填充半刚性线在 18GHz 下的典型损耗为 1.8dB/m。低损耗值使得这类电缆非常适合作为仪器的装置与远程传输。

表 5.2　一些刚性电缆线的近似衰减特性

类　型	在 1GHz 频率下每 30m 的损耗/dB
0.141 英寸半刚性	13
0.5 英寸半刚性	4
0.5 英寸硬线	2.5
0.625 英寸硬线	1.5
1.625 英寸硬线	1

最后要说的是，将一个接口连接到这些电缆的一端时，通常要剥落电缆的部分屏蔽保护层，然后使用电缆的中心导体作为接口的公口针脚。

第6章 无源元件

6.1 引言

本章将讨论射频电路中广泛使用的无源元件的特性。因为当频率达到 GHz 时寄生效应很容易主导电路的行为，因此我们重点关注各种分立元件的寄生电感和电容的简单解析模型。

6.2 射频频率下的互连线：趋肤效应

在低频频率下，我们最关注的互连线特性是电阻率和电流处理能力，也许还有电容。随着频率的提高，我们发现电感可能会变得非常重要。进一步，我们总是会发现电阻由于趋肤效应而增大，就如在第 5 章中讨论过那样。

趋肤效应通常是指随着频率的提高，电流趋向于主要在导体的表面(表肤)流动。由于此时该导体内部区域的载流效率比低频时低，所以导体有用的横截面积减小了，从而使电阻相应加大。

这个既粗略又有些难以理解的描述有可能给读者留下一个印象，即一个导体的所有表面都能同等地承载射频电流。为了更深刻地理解这个现象，我们需要清楚地了解磁场在产生趋肤效应中的作用。为了定性地说明这一点，让我们考虑一个载有时变电流的实心圆柱形导体，如图 6.1 所示。

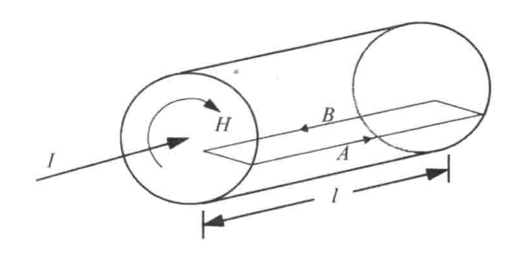


图 6.1　一个圆柱形导体中的趋肤效应的图解

我们暂且假设返回电流(它在任何实际系统中总是存在的)离得很远以至于可以忽略它的影响。时变电流 I 产生了一个时变的磁场 H。根据法拉第感应定律，这个时变磁场会在所示的矩形路径上感应出一个电压。而欧姆定律告诉我们这个感应电压又会在同一条矩形路径上产生一个电流，如图中箭头所示。现在我们可以注意到关键的一点：沿路径 A 的感应电流的方向与沿路径 B 的感应电流的方向相反，因此感应电流要与沿矩形一边的电流相加，却要从矩形另一边的电流中减去。如果仔细注意代数中的正负号，我们可以看到，沿表面的电流加大了，而在表面下的电流减小了。换言之，电流在接近表面处最大，这就是趋肤

效应。

为了更定量地说明这个概念，我们沿这条矩形路径应用基尔霍夫电压定律(需要同时正确考虑感应电压的数值和符号)，由此得到

$$J_B\rho l - J_A\rho l + \frac{\mathrm{d}\phi}{\mathrm{d}t} = 0 \tag{1}$$

式中，J 为电流密度，ρ 为电阻率，而 ϕ 为垂直于所示矩形的磁通量。

正如前面推导的那样，我们可以看到沿着路径 A 的电流密度确实要比沿路径 B 的电流密度大一个数值。这个值会随着深度、频率或磁场强度的增加以及电阻率的减小而加大，这些因素都会加大趋肤效应。此外，式中出现的导数告诉我们电流的变化不只是简单地随深度的增加而减小，而且还会有一个相位的变化。

如果现在使圆柱导体的曲率半径增加到无穷大，我们就把这个圆柱体转变成了常用来分析说明趋肤效应的矩形结构，如图 6.2 所示。我们只是概括性地说明一下这个问题是怎么提出的，然后直接给出解决方案[①]。

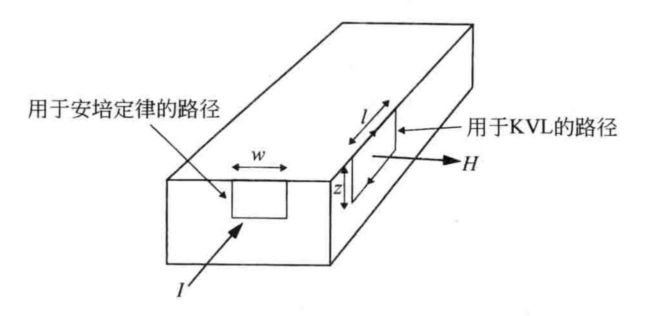


图 6.2　半无穷大导电块的局部截块

计算出由 H 感应出的沿矩形回路线的电压，并像前面那样继续应用基尔霍夫电压定律(KVL)，得到

$$J\rho l - J_0\rho l = \frac{\mathrm{d}\varphi}{\mathrm{d}t} = -\frac{\mathrm{d}}{\mathrm{d}t}\int_0^z Bl\,\mathrm{d}z \tag{2}$$

式中，下标 0 表示在导电块表面处的值。现在我们把 J 和 H(因而还有 B)直接表示成时谐(即正、余弦，译者注)变量，例如，令

$$J_0 = J_{s0}\mathrm{e}^{\mathrm{j}\omega t} \tag{3}$$

式中，下标 s 是任意地选来表示这些趋肤效应变量的幅值。

当代入这些公式时，可以把 KVL 方程写为

$$\rho\frac{\mathrm{d}J_s}{\mathrm{d}z} = -\mathrm{j}\omega B_s = -\mathrm{j}\omega\mu H_s \tag{4}$$

这里利用了磁通密度 B 和磁场强度 H 之间的关系：

① 对于推导的细节，可以参照有关电磁理论的优秀教科书。例如可以参照 S. Ramo，T. van Duzer 和 J. R. Whinnery 撰写的 *Fields and Waves in Communications Electronics*(通信电子学中的场与波)，3rd ed.，Wiley，New York，1994。也可以参照 U. S. Inan 和 A. S. Inan 撰写的 *Electromagnetic Waves*(电磁波)，Prentice-Hall，Englewood Cliffs，NJ，2000.

$$B = \mu H \tag{5}$$

式中，μ 为磁导率(在几乎所有的集成电路中它等于真空的磁导率)。

我们还需要一个方程才能将所需的微分方程组建立起来。安培定律可以引出这个方程②：

$$I_{\mathrm{encl}} = w\int_0^z J\,\mathrm{d}z = wH_0 - wH \tag{6}$$

我们像前面那样进行替换，从而得到：

$$-\frac{\mathrm{d}H_s}{\mathrm{d}z} = J_s \tag{7}$$

联立式(4)和式(7)可以得到关于电流密度的一个简单的二阶微分方程：

$$\frac{\mathrm{d}^2 J_s}{\mathrm{d}z^2} = \frac{\mathrm{j}\omega\mu}{\rho} J_s \tag{8}$$

它的解为

$$J_s = J_{s0}\exp(-z/\delta)\exp(-\mathrm{j}z/\delta) \tag{9}$$

式中

$$\delta = \sqrt{2\rho/\omega\mu} = \sqrt{2/\omega\mu\sigma} \tag{10}$$

称为趋肤深度。可以看到，电流密度从它在表面的值起按指数规律衰减。同样(由第二个指数因子)可以看到，如前所述，这里确实存在一个相移，在深度为 δ 处的相位滞后了一个弧度。

对于这个无限宽、无限长和无限深的导电块来说，趋肤深度就是电流密度减小 e 倍时在导体表面下的距离。铜在 1GHz 时的趋肤深度大约为 2μm。铝的趋肤深度略大一些，约为 2.5μm。这种指数衰减表明使一个导体的厚度超过趋肤深度很多时所能减少的电阻可以忽略不计，这是因为所增加的材料几乎不载流。而且我们可以把它的等效电阻计算成为一个厚度为 δ、电流密度均匀的导体的电阻。这个事实常常用来简化一个导体交流(AC)电阻的计算。为了确保计算结果的合理性，边界条件必须符合用来推导方程组时的那些条件：即返回电流必须在无限远处，并且该导体必须类似于一个半无限大的导电块。只要所有的曲率半径以及各向厚度至少是趋肤深度的 3～4 倍，那么后一个条件就可以很好地得到满足。

作为一个具体的例子，让我们来估算一下一条孤立导线的交流(AC)电阻。假设该导线的直径要比趋肤深度大得多。在这种情况下我们可以假设所有的电流都在一个深度为 δ 的环形区域内流动，由此可以估算出这个电阻值，如图 6.3 所示。这个电阻值很容易计算：

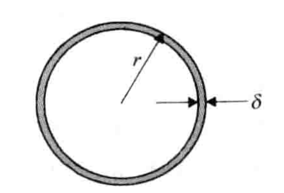


图 6.3　趋肤深度的概念应用到电阻的计算

$$R = \frac{\rho l}{A} \approx \frac{\rho l}{2\pi r\delta} \tag{11}$$

式中，l 是导线的长度，由于我们已假设半径 $r \gg \delta$，因此得到了最终的近似表达式。

在这个情形中，由简单的趋肤深度假设可以得到非常好的结果。但这不是一个典型的

② 回想一下安培定律的文字表述：磁场沿一条封闭路径的积分等于这条路径所包围的总电流。

情形，在许多其他的情况下，这个结果的误差很大。为了估计一下这样做的误差有多大，我们定性地预测一下电流会在何处流动。要估计电流流动的区域，我们回想起在引入趋肤效应时，我们曾基于一些定性的理由，这些理由使我们对几个重要问题有了深刻的理解。这些理由之一就是趋肤效应在磁场最强处也最强。所以，在确定哪些表面可能流过大部分电流时，我们需要识别出什么地方的磁场最强。

考虑一个同轴的导电系统，如电缆，这里有三个表面，但不是所有表面都表现出趋肤效应。外层圆筒形导体流过的是中心导体电流的返回电流。同轴结构是自屏蔽，即电缆外的电场和磁场全为零，这是由于这两个导体的影响被相互抵消了（这个特性就是为什么同轴结构非常重要的理由）。所以，磁场在这两个导体之间的空间中最强，因而在内层导体的外表面和外层导体的内表面上趋肤效应最为明显。在图 6.4 中，电流密度大的区域用黑色来表示。外层导体的外表面几乎没什么电流流过（这里我们再次假设导体的厚度大大超过趋肤深度），因此计算外层导体的电阻时只需考虑它的黑色环形区。注意，并不是所有的表面都显示出趋肤效应。

为了强调“并不是所有表面都显示出趋肤效应”，再举一个例子来加以定性说明。具体来说就是考虑两个距离很近的平行圆柱导体，当其中一个导体中的电流是另一个导体中电流的返回电流时将发生什么情形。与同轴结构中的情形一样，磁场在这两个导体之间的区域上最强，因此最大的电流密度出现在这两条导线互相面对的地方，见图 6.5。不言自明，这就是为什么把电流因邻近导体中有电流流过而集中出现在表面的现象称为邻近效应（proximity effect）的原因。

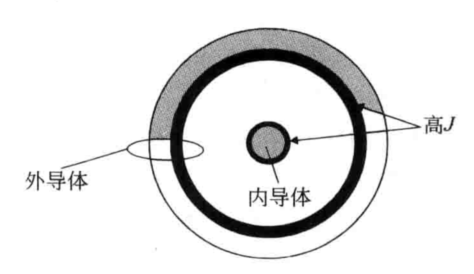


图 6.4 同轴电缆的截面图

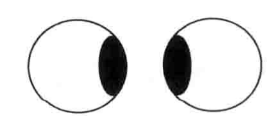


图 6.5 双导线电缆线的截面图

作为最后一个例子，我们考虑一个很薄很宽的导体（同样假设所有其他导体离得非常远）。电流密度大致分布如图 6.6 所示。电流密度向着两端挤塞（为了便于记住这一点，可以把结构想象为一个圆柱导体中心处的一条薄片，这条薄片的两端相应于圆柱导体载流的外表面）。由于沿着这条薄皮两条长边的电流密度较低，因此进一步加宽导体只会使电阻有适度的减小，而加厚导体却会产生很大的影响。

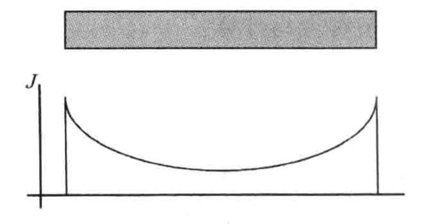


图 6.6 一个薄而宽的导体的截面及电流密度的（近似）分布

如果我们将一个导体放在上面这个导体的附近，那么电流分布情况就会改变，从而使一个本来就复杂的问题变成一个几乎无法求解的。

计算后两种结构的等效电阻显然不像计算孤立导

体时那样简单。确实，对看似简单的一个单层绕组的结构精确计算其有效电阻已经实际上是不可能的，因为在线匝之间存在着那么多相互作用。这个困难更进一步说明了自动假设所有表面都能有效地载流是很有风险的。

最后，为得到一系列有关趋肤效应的有关公式，可参见 Harold A. Wheeler 发表的 "Formulas for the Skin Effect"（有关趋肤效应的公式，*Proc. IRE*，v. 30，September 1942，pp. 412-24）一文。

6.3　电阻器

在高频下，即使像电阻这么简单的元件也会表现出很复杂的特性。如果我们首先意识到电流流经连接引线和电阻两者，可以构建一个非常简单的电阻模型。与电流相关的储存在磁场中的能量说明有某些串联电感的存在（作为一个非常粗略的近似，在轴向封装中的引出脚导线寄生电感大约 0.5nH/mm）[3]。另外，由于导体两个端口间有一段距离，因而必定有电容与这个电阻并联。电阻的最简单的（但不会是唯一的）RF 集总电路模型看起来如图 6.7 所示。

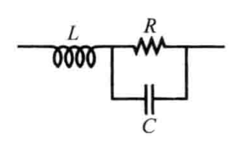


图 6.7　简单的集总参数射频电阻模型

寄生电容和电感的存在使得阻抗不再是一个与频率无关的纯电阻。对于电阻值非常低的电阻器，随着频率的增加，起始时阻抗是增加的，这发生在大约当串联电感的电抗与电阻值相比是不能忽略时的频率下。类似地，高值电阻则由于并联电容的存在而发生过早的（随频率的）阻抗下降。对于某个具有中间（确定）电阻值的电阻器，它们阻抗值大约保持不变的频率范围可能会是最大（至少对于我们这个简单的模型而言）。根据传输线理论，可以得知这个奇妙的值为

$$R_{opt} = \sqrt{L/C} = Z_0 \tag{12}$$

我们不在此进行方程(12)的推导，而先写出阻抗大小的表达式然后求解最大平坦化的条件。这个条件表述为要有尽可能多的导数在频率为零时其值为零[4]。通过考虑两种极端的情形，我们可以构造得到式(12)的一个直观路径。假如电感被忽略，网络的时间常数就是 RC。假如电容被忽略，那时间常数是 L/R。令这两个时间常数相等解得的 R 即为表达式(12)。如果电阻值比 Z_0 小，则等效电路中的阻抗只会随频率增大而增大，而且其阻抗值与用弧度为单位的频率曲线上是拐角频率（corner frequency）由 R/L 近似给出。如果电阻值比 Z_0 大，则阻抗起初降低，其拐角频率（频率用弧度表示）近似等于 $1/RC$。如果电阻值等于 Z_0，阻抗值近似为常数的频率带宽 f_{max} 由谐振频率 LC 给出。LC 越小，这个电阻器呈现电阻特性的对应频率范围就越大。

作为参考，表 6.1 中给出了一些有代表性的电阻值（近似值）。需要注意的是对应最大平坦化的阻抗为 100～200Ω。这也是传输线阻抗值趋向于在这个范围内的原因。

---

③　更多电感和电容的方程可以参考：T. Lee, The Design of CMOS Radio-Frequency Integrated Circuits（CMOS 射频集成电路设计），2rd ed.，Cambridge University Press，2004.

④　一如前面提到过的，对阻抗的平方来进行这个步骤通常更容易些。

表 6.1　简单的集总参数射频电阻模型的近似元件值

电 阻 类 型	L	C	Z_0	f_{max}
0.5W 轴向引线(总长度 3cm)	15nH	0.5pF	170Ω	1.8GHz
0.5W 轴向引线(总长度截短到 1cm)	5nH	0.5pF	100Ω	3GHz
0.25W 轴向引线(总长度 2cm)	10nH	0.25pF	200Ω	3GHz
0.25W 轴向引线(总长度截短到 0.7cm)	3nH	0.25pF	100Ω	6GHz
型号 0805 表面贴片	<1nH	0.02～0.1pF(见正文)	150Ω	20GHz

再次说明，表中所列数值为估计值，只能作为参考。特别是贴片电阻中的并联电容很难保持是常数，它的值和其结构、材料构成等有很大关系。在大多数情况下，测量到的值与其固定的环境和方式的关联度比贴片电阻本身的电阻值更大。从测量数据中提取模型参数时如果没有考虑这点，那么提取出的贴片电阻的电容值往往可能跨越几个数量级。作为一个例子，在微带线环境下一个贴片电阻特性与在其他种类的封装结构下的特性就不一样，这是因为接地面的不同会改变电场的分布。在这种情况下，图 6.8 所示的等效电路模型就更合理些。

对 0805 型号的电阻，其典型的模型参数值可能为 C=80fF 和 L=0.7nH。需要注意的是此模型中没有直接并联在电阻上的电容。可能会有电容并联在电阻上，但这个电容值是如此之小，不将其包括在内依然可以得到与测量结果十分吻合的拟合。要记住的重要的一点是并不存在唯一一个"正确"的模型，许多不同的模型都能得到可接受的结果。例如，在许多 CAD 工具中使用的模型经常是如图 6.7 和图 6.8 模型结合的模型结构，如图 6.9 所示。

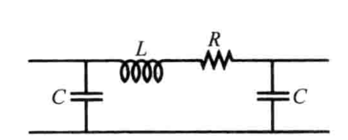


图 6.8　在一个微带线环境下的表面贴片电阻的更完整的模型

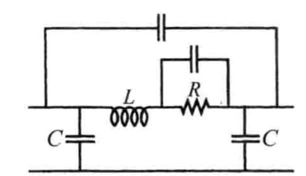


图 6.9　表面贴片电阻的另一个模型

注意：最优的 Z_0 值都超过了典型的传输线阻抗(例如 50Ω)。假如要给一个 50Ω 的传输线提供终端负载，那么最大带宽是由两个电阻器的并联组合得到而非一个单独的 50Ω 电阻器。在带宽超过 1GHz 的时候，四个 0.25W，每个电阻阻值为 200Ω 的电阻并联提供 50W 终端负载是一个更好的结构。类似地，在带宽超过 10GHz 时两个 0805 贴片电阻的并联组合也可以作为极佳的终端负载使用。表 6.1 也说明了为什么传统的电阻(特别是 0.5W 类型)很少在微波电路中使用。如果需要更大功率的终端负载，我们偏向于使用并联低功率、高频率的电阻，这样可以得到比直接使用高功率的电阻更宽的频带。

作为一个补充说明，表面封装的贴片元件，其数字标识符指明了元件一些物理尺寸。在截断误差范围内，头两个数字代表元件以 mm 为单位的长度的四倍，第二对数字则是其宽度的四倍⑤。表 6.2 所示为一些常用的尺寸及其功耗级别。

⑤　最初的规定是以 mils(密耳)来计算的(一英寸的千分之一)。因此，每对数字都要乘以 10 来得到元件以 mils 为单位的近似尺寸。

表 6.2 一些表面贴片电阻的封装与特性

封装	大致尺寸/(mm×mm)	P_{diss}(功耗)/mW
0402	1.0×0.5	60
0603	1.6×0.8	60
0805	2.0×1.25	100
1206	3.2×1.6	125
1210	3.2×2.5	250
1812	4.5×3.2	500
2512	6.4×3.2	1000

寄生电容在某种程度上取决于元件的一端是否接地、组成元件的主要部分的材料以及元件是否完全安装在一个接地面上或元件终端有一端悬空。因此，我们提供不出对所有这些情况都成立的寄生电容的一般正确值。尽管如此，我们仍然能给出与这些封装相关的寄生参数的大致信息。考虑到在表 6.2 中所列的最大尺寸的电阻，2512，其 6.4×3.2×3.2 的尺寸具有寄生电感大约为 2.5nH，其典型的并联电容值为 0.18pF。如果测量值不具备的话(经常是这样子的)，可以借助于下面这个计算无穷薄平板电感的公式来估计其他封装件的串联电感

$$L_{sheet} \approx \frac{\mu l}{2\pi}\left[0.5\ln\left(\frac{2l}{w}\right)+\frac{w}{3l}\right] \tag{13}$$

由于电阻性的材料一般总是平的，淀积在较厚的绝缘衬底表面(甚至是“厚膜”电阻器)的薄层，因此这个公式是合适的。对于在低功耗贴片元件上通常遇到的宽长比，电感一般是在 0.3～0.5nH/mm 范围内。

我们可以两次使用这个方程来估算总的寄生电感。总电感的一个部分是对应于主体的电感，因此其长度和宽度首先代入方程。对这项(通常)是占主要成分的项，我们还必须加上由于平坦的元件的垂直端面引起的电感。只不过现在要用其高度来替代方程(13)中的宽度。然而由于半月形的焊接面加厚了这些垂直面，这里不宜使用各个垂直面的全部电感。作一个粗略的选择，取计算值的 1/2～2/3 是一个不错的因子。用 1/2 这个因子，我们估算出 2512 型封装件的电感约为 2.6～2.7nH，这个值与测量数据吻合得不错。

需要注意的是这里的单位长度的电感比 1nH/mm 的估算值要低，这些估算通常值适用于薄的、圆头的元件的引线。这是因为又厚又宽的贴片组件会分散磁场线，减少了磁通密度进而减小了电感。

如前面解释过的，估算这种结构组件的电容比较复杂。但是，我们可以给出一个基于圆柱型导体构成的偶极子天线的单位长度电容表达式(推导在第 21 章给出)的粗略的对电容值下限的估算公式：

$$C \approx \left\{c^2\frac{\mu_0}{2\pi}\left[\ln\left(\frac{2l}{r}\right)-0.75\right]\right\}^{-1} \approx \frac{2\pi\varepsilon_0}{\ln(2l/r)-0.75} \approx \frac{5.56\times10^{-11}}{\ln(2l/r)-0.75} \tag{14}$$

这个公式给出一个在地与长为 l、半径为 r 的导体之间的电容(单位长度)的估算值。对于典型的贴片组件的尺寸而言，这样给出的值一般在 15fF/mm 乘以或除以因子 2 之内。这个公式实际上假设了在一个与接地面垂直的而在其面上的圆柱导体，其中一端接地(其中 c 为光

速，详见第 2 章公式(12-14)，译者注)。

由这个公式计算得到的电容值的下限假设了(还有其他的条件)单位介电常数以及没有附加的金属焊盘，等等。由于封装件和 FR4 的介质常数的影响，其电容值会升高。而且，PC 板和电阻终端之间不连续过渡也会引起在每个元件端口与地之间的并联电容。即使贴片元件的这些特性很难定性地说是唯一的，但是它们本身的小电容值意味着相较其他离散电阻来说，这个效应更为明显。

这个近似的方法可以用来估算贴片电感的封装寄生电容，以及电路中其横截面是圆的元件的寄生电容。只是需要记住，封装件寄生参数只是整体寄生的一部分，一些寄生可能在元件的内部产生(例如，电感中的圈与圈之间的电容)。这又一次表明，计算出来的封装电容仅为寄生电容的下限的估计。

作为最后的总结，需要注意许多轴向引线的电阻主要是由碳构成的，这些阻性粉末形成了圆柱形引线。不幸的是这种电阻会展现严重的 $1/f$ 噪声，因为其功率谱密度与通过电阻的 DC 偏置电流成正比。碳膜电阻在这方面性能要好得多，而金属层电阻则更好。虽然 $1/f$ 拐角频率远在 RF 范围之下，但我们必须知道振荡器可以把低频噪声上变频至载频附近的相位噪声[⑥]。因此，尽管 $1/f$ 噪声在诸如射频放大器的电路中通常不是一个大问题，但它在所有的射频电路中都是不能被忽略的。可庆幸的是，贴片电阻则由各种薄膜做成(因而，不受碳阻上述特性的影响，译者注)。

6.4 电容器

许多不同的电介质材料被用来制作电容，以尽可能满足对电容的许多相互矛盾的要求。击穿电压、温度系数、RF 损耗以及电容密度之间的折中选择不可避免地产生了许多种目前可提供的电容。限于篇幅，这里不可能罗列所有种类的电容，因此只集中讨论在高频电路中通常碰到的电容。

损耗最低的电容是把空气(或真空)作为电介质。高密度、低损耗的更高密度的电容可以利用云母(一种自然产生的矿物)和聚苯乙烯。尽管聚苯乙烯有极佳的电特性，但它的熔点过低，这就限制了其只能在低于 85℃的环境下使用。因此在焊接聚苯乙烯电容时必须格外小心。

用 PTFE 材料做成的电容要耐用得多，这种材料(就如我们先前提及的)也是一种极低损耗的电介质。但是，制造高质量 PTFE 薄膜的成本也意味着用此材料制作的电容有相当厚的介质层厚度，这就导致了低的电容密度(但是有非常高的击穿电压)。

陶瓷电容本身则可以有很多种类，它们根据其用到的介质特性来区分。为了便于追踪那么多的排列组合，电子工业协会制订了由三个词构成的命名方法。第一个字(第二个)是一个字母，表示最低(最高)工作温度，第三个字也是一个字母，指出在整个工作温度范围内电容值的最大波动。表 6.3 列出了一些具体的型号。例如，标号为 X7R 的电容在工作温度 $-55 \sim +125$℃的范围内，其电容值变动范围最多为±15%。

⑥ Lee，前面引用过的(见脚注 3)。

表 6.3 三字母的电容器码(EIA)

Temperature/℃		在温度变化区间内电容变化的最大百分比	
最小值	最大值		
X：−55	3：+45	A：±1	P：±10
Y：−30	4：+65	B：±1.5	R：±15
Z：+10	5：+85	C：±2.2	S：±22
	6：+105	D：±3.3	T：−33，+22
	7：+125	E：±4.7	T：−56，+22
		F：±7.5	T：−82，+22

虽然温度系数为零是最希望达到的目标，但在有些情况下设计人员使用具有特定 TC 值的电容替代零 TC 电容。比如在振荡器中，电感一般呈现一个正的 TC 值[⑦]，因此电容必须采取一个负的 TC 补偿值来达到振荡器整体的 TC 值为零。可控温度系数的电容一般用字母 N(表示负的)或 P(表示正的)来标识，而且还有以单位为百万分之一每度计量的最大 TC 值。例如 N750 就表示此电容温度系数为−750ppm/℃。[⑧] 使事情变得复杂的是，还有另一套标识方法来传递同样的信息。为了节省在小的器件上打印标记所需的空间，专门设计了一个三位数字的 EIA 码，但这个码并不能直接传递有关实际的 TC 的数值信息，因此在表 6.4 中给出了两种标设方式的转换信息。

表 6.4 电容器的 TC(温度系数)码

老的表示法	3 位数 EIA 码	老的表示法	3 位数 EIA 码
NP0	C0G	N330	S2H
N033	S1G	N470	T2H
N075	U1G	N750	U2J
N150	P2G	N1500	P3K
N220	R2G	N2200	R3L

三位的 TC 码规则中，第一个字母表示 TC 的高位数，其数值是标准电阻值的一个子集。例如，我们从表中可以看出来 $P=1.5$，$R=2.2$，$S=3.3$，$T=4.7$，$U=7.5$。中间一位数表示指数幂。标识 NP0(C0G)表示“负-正-零”，意思是其组成材料中由正的 TC 值和负的 TC 值产生一个零 TC 值(典型的最大为±30ppm/℃)[⑨]。因此电容值在军用温度范围−55～+125℃内变动在 0.15%之内。在标准的贴片封装器件中高达 10nF 的电容也是可用的。在标准类型中 NP0/C0G 的损耗为最低，其 Q 值在低频的时候也能在 500～600 之间，而且具有低的温度系数。

其他常用的材料包括稳定性稍差(但是有更高介质常数)的 X7R 陶瓷，使用此材料的贴片器件其值可达到大约 100nF。正如前面提到的，电容在军用温度范围内波动可能会达到±15%。与 C0G 不同的是，当 DC 偏压增加时此材料的电容值下降(呈近似线性关系)，在

⑦ 考虑到电感在尺度上和长度是成比例的，而且大部分材料受热都会膨胀，因此大部分物理实现的电感的 TC 值都为正的。

⑧ 1ppm=10^{-6}

⑨ 注意这里标记符为数字 0 而非字母 O。

额定电压下可能会有额外的30%的下降。随着偏压变化的电容值与介质的压电性有关，在信号通道中，这种非线性行为可能产生非常大的失真。（另外，X7R材料的（电容）损耗是C0G的两个数量级。）

高*K*（高介电常数）陶瓷，例如Y5V，可以制成体积最小的电容但其TC值却极高（例如在零偏置下，温度范围－30～85℃内电容值会下降令人吃惊的80%），以及高达X7R的1/3损耗。其电压系数也是极其大的负数，在额定电压下电容值下降高达75%。这种类型的电容可以用作有效的混频器，因此要注意（或者说可以利用这个特点）。而且，这种介质材料带有强压电性，经常会发生因为剧烈的机械冲击而致使电压尖峰的出现（有时可高达数十伏）。即使这些尖峰电压不会引起对敏感电路的直接损伤，但应该注意的是，高*K*介质电容器的麦克风行为会导致一系列不希望发生的问题，特别是当它们连接到一个电路的敏感节点、并且处在振动状态（如移动的应用）。因此这些电容最常用于电源的旁路电路中，而不是在信号通道。在标准的贴片封装器件中，电容值可高达1μF。

我们也可以把PC板作为电容的介质层，将所要求的电容值的一部分用PC板实现对于在已经制作完毕的电路后调整电容值的目的，显得非常方便。无论在什么情况，知道有多少电容值是与一个给定面积导体相关联是一个好的主意，即使仅仅是为了估算版图的寄生参数。对于FR4材料，我们可以在1/32″(0.8mm)厚的衬底上制作5pF/cm^{2}的电容或1/16英寸(1.6mm)厚的衬底上制作2.5pF/cm^{2}的电容。FR4的损耗是可以接受的，比起X7R和Y5V要好一些。当然，更低的损耗（有时也是更小的电容）也可以用高品质的PC板材料，如PTFE或RO4003制作。更多关于使用PC板走线作为电容电感的实现的讨论在第7章中进行。

电容值被编码成三位数字，打印在电容器上（如果数字放得下的话），后面还有一位字母表示其误差（见表6.5）。标注的前两位是用整数表示的有效数字，后一位是指数幂。其单位是pF。所以，“221K”就表示一个容值为220pF(22E1 pF，译者注)、偏差为±10%的电容，“105M”就表示一个容值为1mF(10E5 pF，译者注)、偏差为±20%的电容。偶尔也会使用其他的标识法，但这种方法是使用最为广泛的。如果对标注值有所怀疑，总是可以通过实际测量来验证。

表6.5　电容器的容差码(EIA)

标识符	容差/pF	容差/%
B	±0.1	
C	±0.25	
D	±0.5	
E		±25
F	±1	±1
H		±2
I		±5
K		±10
M		±20

就像与电阻的情况一样，在射频频率下电容的寄生效应不能被忽略。最为简单的实际电容器的集总参数射频模型包括引线或终端电感（如前所述，典型的圆导线引线的电感值约

为0.5～1.0nH/mm)，以及模拟损耗的电阻项，见图6.10。对于贴片封装器件的电感可以用前面的公式(13)来估算。

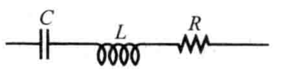


图6.10　简单的集总参数电容器模型

这个模型中的电阻项用来考虑至少两种不同的机理。一种是介质材料的损耗，其他是导体的损耗(该损耗在高频时因趋肤效应加剧)。损耗通常用损耗因子 D(或者用损耗角正切值 $\tan\delta$)来表征。损耗因子就是 Q 值的倒数，而损耗角正切值定义为介质常数的虚部和实部的比值。严格来说，损耗角正切值只适用于电介质材料，但常常被用来包括了电容中所有的损耗。在后一种情况下，损耗角正切值和电容损耗因子是一样的。不同描述损耗的方法是由习惯而来的。电力电子学中倾向于考虑功率因子(即电压与电流相位角的余弦，此角度和阻抗角相同)，射频工程师则习惯于用 Q 值来描述问题，而材料科学家却倾向于把注意力放在损耗角正切值(或损耗因子 D)上。

功率因子的定义意味着其与 Q 的反正切角的余弦是相同的(证明过程留给读者去完成，因为很明显你不会对这个推导过程有多大兴趣)。对于充分大的 Q 值，功率因子和损耗因子趋于相同。例如，Q 值超过7时就可以确保低于1%的误差。对于所有在信号通道中实际使用的电容来说，Q 值一定会大于这个值，因此在实际中可以将功率因子和损耗角正切值视作相同。

基于这些定义，由于电介质损耗引起的有效串联电阻(ESR)这个分量是：

$$R \approx D/\omega C \tag{15}$$

这个表达式只有在频率远低于串联谐振频率时才适用。很清楚，ESR是个与频率相关的量，特别是当计及趋肤效应的导体损耗也被考虑到时。

当频率远高于谐振频率时，因为电感性电抗占主导地位，所以电阻变得与频率成比例关系，因而有下面这个近似表达式：

$$R \approx D\omega L \tag{16}$$

我们可以从上面这个 RLC 串联电路中推断出一些重要的信息。高于网络的谐振频率时，电路总体是呈感性的，因此阻抗随着频率的增加也在增加。在达到谐振频率时阻抗值最小。如果使用了一个电容作为电源的交流旁路支路，应该注意到在较高频率时因为串联电感的影响这个旁路电容所起的作用会降低。然而，只是因为呈现出电感性，这个电路不应该妨碍其使用，因为最关心的量是阻抗的幅度。如果这个阻抗值是足够低的话，即使是工作频率高于谐振频率，这个电容仍然可以当做旁路元件来使用。

作为对这些寄生元件值的一个大致的了解，表6.6中列出了(100MHz时)参数值。假定一个圆片电容的总长度(引线两端，包括电容体之间的距离)大约为10mm。这个100MHz的测试频率大大地超过了这个圆片电容的17MHz的自谐振频率，因此有效串联电阻更多的是由引线电感引起的而非本征电容。通过缩短引线长度，我们有可能增加自谐振频率并且在一定程度上减小 R 值。

表6.6　有代表性的电容器及在100MHz下的集总模型参数

型　号	C	L	R	SRF
陶瓷圆片(C0G/NP0)	10nF	10nH	0.5Ω	17MHz
0805 C0G/NP0	10nF	～1nH	0.08Ω	50MHz
0805 C0G/NP0	100pF	～1nH	0.25Ω	500MHz

需要再次说明的是，损耗是相当地依赖于频率与介质材料的函数，因此表 6.6 中的阻值不能被认为是一个普遍成立的常数。其值在不同的情况下会有所不同。

6.5 电感器

6.5.1 表面贴片电感和铁氧磁珠

对小型化电路的永无止境的追求导致产生了包括电感在内的很多微小的元件。适用于射频与微波应用的典型的贴片电感值从 10nH 到 1μH 都有产品。由于制造商一般不怎么提供这些电感的详细信息，使用者在使用前有必要对其进行实验测试。尽管这么说，我们可以提供一些大概的一般信息以便进行初步的计算。大部分贴片电感的并联电容都分布在很小一个范围内，其值一般为 0.1～0.2pF，这对于最小的电感值，其自谐振频率相当于 4～5GHz，而对最大的电感值，则为几百兆赫兹。假如需要更高的自谐振频率，那就得自己做电感了。同样，当需要一个比可买到的功耗更低的电感时，你也只能靠自己来做。

偶然有时候，需要增加一个电路的射频损耗。比如说，这种需要可能会因要抑制一个放大器的痼疾似的寄生振荡而产生。对这个场合，在晶体管基极（或栅极）加一个串联的“破坏 Q 值”的元件一直以来是十分成功的（的确，这个诀窍可以追溯至真空管时代的早期）。这些烦人的振荡经常发生在一个频率点，此时收集极纯负载看起来是电感性的，而晶体管的增益依然很大。因为存在着通过收集极-基极电容的反馈，这个感性的负载可以产生负的输入阻抗。有了电感、电容及负阻的这个组合，振荡必然要发生。在基极串联一个足够大的电阻可以有效地确保一个基极纯电阻为正，但其缺点是会引起一个 DC 的电压降。

为了引入仅在射频下的有选择性的损耗，可以采用铁氧磁珠。它们是由有损磁性材料构成的圆柱形元件。将导线穿过珠子的中心一次或绕两次（有需要的话，可以更多）就构成了一个变压器，它的次级负载为代表有损铁氧体的电阻，见图 6.1。因为在射频下才有耦合，因此这里没有直流损耗。

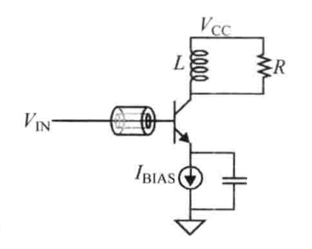


图 6.11 铁氧磁珠的典型用法

铁氧磁珠在抗电磁干扰中使用十分广泛（EMI，electromagnetic interference，也称为 RFI，radio frequency interference，即射频干扰）。它们能被用到如此众多的场合，以至于许多工程师会不切实际地指望这种简单的元件有太多的应用。

6.5.2 电感公式

我们已经在很多前面的方程中零散地用到了电感。现在我们要给出另一些适用于通常会遇到几何尺寸的电感表达式。在下述情况下，方程只适用于直流情况，除非是特殊声明。在高频下，趋肤深度的减小会使得电感下降，这是因为导线内部的磁通量消失了。幸运的是，内部的磁通量只占总体的一小部分（<5%），因此它的降低对整体电感不会造成很大影响。然而，通过明确地知道在公式推导时，哪些假设是做了，可以帮助我们避免不必要的惊异。

1. 平坦薄层导体

我们已经给出了一个计算电流薄层的电感公式。这里重新写出

$$L_{\text{sheet}} \approx \frac{\mu l}{2\pi}\left[0.5\ln\left(\frac{2l}{w}\right)+\frac{w}{3l}\right]=(2\times10^{-7})l\left[0.5\ln\left(\frac{2l}{w}\right)+\frac{w}{3l}\right] \tag{17}$$

2. 导线电感

人们经常想知道一定长度的导线电感，这或者是因为要定量地计算寄生电感，或者是需要把电感用作为一个电路元件。如果可以忽略附近导线的影响（即假设回流电流在无限远处），则一个截面为圆（半径 r，译者注）的导线的直流电感可以由下式计算[10]：

$$L \approx \frac{\mu_0 l}{2\pi}\left[\ln\left(\frac{2l}{r}\right)-0.75\right]=(2\times10^{-7})l\left[\ln\left(\frac{2l}{r}\right)-0.75\right] \tag{18}$$

对于一根 2mm 长的用作标准集成电路的键合线，这个表达式给出的电感为 2.00nH，这就导致了这样一个估算法则：一根细的圆导线的电感大约为 1 nH/mm。注意，导线的电感随长度的增长要比线性增加来得快，这是因为在导线的各部分间有相互耦合（即，这里有弱的变压器作用），且其极性是帮助电感增加的。不过从对数项来看的话这个效应并不大。例如，导线的长度从 5mm 增加到 10mm 时，每毫米的直流电感值从 1.19nH 变化到 1.33nH（至少根据式(18)计算的是如此）。同样的，导线直径的变化对电感影响不大，因此甚至在分立电路中能遇到的更大导体的电感值也保持在同样的数量级（例如 0.5nH/mm）。

如果附近有一个导电平面——可以松散地定义为两者之间距离大约接近导线的长度——则电感将明显比由式(18)计算出的要小。直觉上来讲，这个电感的降低可以用以下方式解释。流经导线的电流（可以想象为正电荷沿 X 方向移动）在接地平面上感应出了一个镜像电流（比如说，负电荷沿 X 方向移动）。极性相反的电荷沿同一个方向移动相当于两个电流在不同方向上流动，因此它们的磁场会相互抵消，导致磁通量的减少。导体平面越接近，磁通量的减少就越严重（因此电感减少也越多）。

3. 空心螺线管

虽然我们关注的元件都是可以用平面工艺实现的，更多的单位体积的电感也可以用教科书上描述的经典 3D 结构获得：即单层螺线管，如图 6.12 所示。

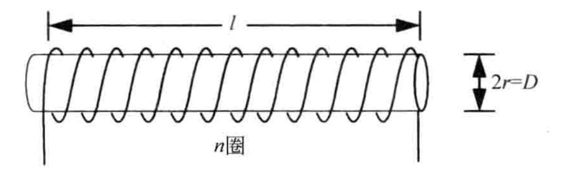


图 6.12　单层螺线管

假设线圈不像图中那样，是紧密排布的（“密布”），则以 mH 为单位的电感值可由下面这个由 Wheeler 于 20 世纪 20 年代晚期提出的著名的公式计算

$$L \approx \frac{n^2 r^2}{9r+10l} \tag{19}$$

⑩　ARRL 手册，美国无线电传接联盟，1992，pp. 2-18。当导体平面之间相距很远时（至少相距导体长度的 1～2 倍），它们间的邻近效应可以被忽略。

这里 r 和 l 是以英寸为单位的[11]。在 SI 单位中，公式为

$$L \approx \frac{\mu_0 n^2 \pi r^2}{l + 0.9r} \tag{20}$$

这里假设了自由空间磁导率。当长度超出半径的三分之二时，这些公式对紧排、单层的线圈给出了令人惊讶的精度（一般比 1%还要好）[12]。这个公式的结果相对于 $l=0.4r$ 的实际电感是低估了大约 4%，因此该表达式对于比较短的线圈也是有用的。

对于那些对这个著名而又被普遍使用的公式起源感兴趣的读者，它的推导始于标准的大学物理中有关无限长螺线管方程。对于这个无限结构的任何一段长为 l 的导线，其在 SI 单位下的电感公式为

$$L = \mu_0 n^2 \frac{A}{l} = \frac{\mu_0 n^2 \pi r^2}{l} \tag{21}$$

这里 A 为螺线管的横截面积，n 为在这一段长度内的线圈匝数。需要注意的要点是，在其他参数保持不变的情况下，电感值会随着螺线管长度的增加而下降。

沿着一个有限长的螺线管的磁场强度在其两个末端自然会减小，因此电感也会降低，螺线管电学长度要比其物理长度长。如果螺线管的长度比半径要大好多，这个有限长度的效应不会被那么明显地感受到。因此，对于末端效应的纠正是长度与半径比率的函数。Nagaoka[13] 在一篇著名文章中给出了这个纠正因子的一张表格和一些曲线，由 Grover 所做的其后工作又提供了一个无限序列，这个序列可以根据要达到的精度来截断。从他们的工作中我们可以看出，作为第一次近似，对有效电学长度的一个简单的估计就是物理长度，加上半径。这种预设的纠正方法与对电容中边缘项做法相似，而这样做的结果也的确非常理想：

$$L = \frac{\mu_0 n^2 \pi r^2}{l + r} \tag{22}$$

这个公式是非常准确的（而且相对来说比较容易记忆，特别是使用 SI 单位制时），但是一个更严密的分析表明它的确稍微低估了电感值，特别是对短线圈的电感。Wheeler 的公式通过增加 90%而不是 100%的半径加到长度上，达到了更高的精度。

与电感的两个终端相并联的有效电容在很大程度上依赖于边界条件。例如，假如一端是接地的，则有效电容对邻近线圈间的电容不怎么敏感，而是更多地取决于外部边缘到地的电容。后面的这个电容是难以从解析公式计算的。据作者所知，迄今还没有发表过通用的解析解[14]。因此，我们这里能提供的只有半经验公式，这里假设了：(a)导线间的绝缘，其相对介质常数为 1；(b)一端是接地的。在这些假设成立的时候，有效的并联电容可以由下面公式近似计算得到[15]：

---

[11] H. A. Wheeler, "Simple Inductance Formulas for Radio Coils"（无线电线圈的简单电感公式），*Proc. IRE*, v. 16, no. 10, October 1928, pp. 1398-1400.

[12] 如后面会讨论的，最高 Q 值一般来说是对某些特定的导线直径与线间距的比率来得到，而其准确值则与线圈长度和线圈直径比有关。

[13] "The Inductance Coefficients of Solenoids"（螺线圈的电感系数），J. College of Science (Tokyo), v. 17, art. 6, 15 August 1909, p. 1.

[14] 虽然发表过很多种算法，但是仔细一看，都有很大的误差。

[15] 这个方程是基于 Medhurst, *Wireless Engineer*（无线工程师，February 1947, pp. 35-43; March 1947, pp. 80-92），书中的数据。其中的系数已经被选择用来改进 Medhurst 提出的公式的精确度和减少其复杂性了，以及为了采用 SI 单位制。

$$C_{\text{coil}} \approx \pi\varepsilon_0\left[0.4(l/D+1)+0.9\sqrt{D/l}\right]D \tag{23}$$

这里 l 和 D 分别为线圈的长度和直径。这个公式在 l/D 值从 0.1～50 范围内和 Medhrust 的数据相比，误差小于 5%（大约与其数据的误差相当）[16]。需要注意的是主要的依赖关系是在线圈的直径，而对总长度的依赖关系则较弱。因此，对于一个给定数值的电感，当其半径最小时可获得最高的自谐振频率。

为了得到准确估算的总并联电容，必须要注意仔细地包括任何长度的未接地引线的电容。为了达到此目标，我们可以使用对一根孤立导线的单位长度的电容计算公式（为了查看方便起见，我们这里给予重复）：

$$C_{\text{wire}} \approx \frac{2\pi\varepsilon_0}{\ln(2l/r)-0.75} \approx \frac{5.56\times10^{-11}}{\ln(2l/r)-0.75} \tag{24}$$

回想起，典型的值一般数量级为 10 pF/m。

除了电感值和寄生并联电容外，有效串联电阻是十分重要的。为了估算它，人们十分自然地会想到使用趋肤效应的公式。不幸的是，趋肤效应公式的得到是基于经受均匀磁场的一个半无限大的导体块。但在螺旋管中情况却大不一样：一个线圈的磁场会扰动流经邻近（及其他）线圈中的电流分布，因此边界条件（而因此，有效横截面）发生了显著的变化。使用非经修正的趋肤效应公式因此通常会产生相当大的误差。如果想要处理普遍的情况，用解析公式估算损耗是困难的，但如果我们只关注导致最大 Q 值的条件，则可以有相对简单的表达式。

Medhurst 的深入研究使得我们可以简洁地表示取得空心电感的最大 Q 值的一些条件。在下面的讨论中，我们理解到这些优化条件相对来说是比较平坦的，因此得到的 Q 值不会对这些优化条件的偏离十分敏感。

首先，Medhurst 用一个简单得不得了的表达式给出了线圈的 Q 值：

$$Q = 7.5D\psi\sqrt{f} \tag{25}$$

D 是以米为单位的线圈直径，f 是频率（Hz），ψ 是长度 l、线圈直径 D 以及导线直径 d、绕组线圈之间的间距 s 的一个复杂函数。频率的平方根项是基于趋肤效应的考虑（假设工作频率远低于自谐振频率），但这就是仅有的我们可以猜测的知识。据作者对于除铜之外的金属来说，必须使用材料的相对电导率来缩放常数 7.5。

函数 ψ 很难以简单的解析形式表达出来，因为其与 l/D 和 d/s 都有关系。不过，如果我们的目标只是想达到最大 Q 值，那么在每个 l/D 下得到最大 ψ 值的表达式也是不难的。这样的一个式子对于分析是没有太多帮助的，但很适合于设计时使用：

$$\psi_{\text{opt}} \approx 0.96\tanh(0.86\sqrt{l/D}) \tag{26}$$

这看起来是一个经验公式，但是这个表达式能在 l/D 高达 10 的情况下以好于 3.6% 的精度拟合 Medhurst 的数据，在 l/D 大于 10 时精度也胜于 4.5%。需要注意的是最优 ψ 值仅仅是 l/D 的函数。选取 l/D 值大约为 5 时可得到 ψ 为 0.88，这已经极为接近在所有情况下最大的 ψ 值了，采用更大 l/D 的比值并不能帮助得到更高的 ψ 值。

由式(26)隐指最大化 Q 值要求同时最大化 D 和 ψ，必须选择这样的最大线圈直径以至

[16] Medhurst 声称他的公式有更高的准确度，但实际上他的最高误差有 8% 那么大。

于使得自谐振频率远高于设定的工作频率(例如，至少 2～3 倍)。人们可以与要求的电感值一起使用式(23)来计算 D 的上限。

已经有效地计算好 D 和 l(l 则是由选好的 l/D 决定的)，我们可以使用 Wheeler 的方程来计算线圈的匝数 n：

$$n=\frac{\sqrt{L(0.9r+l)/(\mu_0\pi)}}{r} \tag{27}$$

绕组的密度(pitch)则是每米 n/l 圈数。

导线直径与绕组线圈间间距(这个间距由相邻导线的中心间测量得到)的比值可以粗略估计为

$$\frac{d_{\text{opt}}}{s_{\text{opt}}}\approx\begin{cases}0.5+0.07(l/D), & l/D\leqslant 7 \quad (28)\\ 1, & l/D>7 \quad (29)\end{cases}$$

再次说明，最后这两个方程是从粗略拟合 Medhurst 的数据得到的，且给出了 d/s 的值，这些值导致表达式(26)的成立。

最优的导线直径因此为

$$d_{\text{opt}}\approx\frac{l}{n}\left[1+\frac{s_{\text{opt}}}{d_{\text{opt}}}\right]^{-1} \tag{30}$$

绕组线圈间间距(这个间距由相邻导线的中心间测量得到)通过将表达式(28)或(29)代入表达式(30)中的相应比值而得到。

设计这些线圈时最后一点有用的信息是关系到导线的性质。纯铜的电导率是大约 5.7×10^7 S/m。光铜线的直径一般是以表格的形式给出，但简单的(虽然是近似的)计算公式为

$$D\approx\frac{0.32}{10^{(\text{AWG})/20}} \tag{31}$$

这里 D 是以英寸来计量的，AWG 是美国线径标准(American wire gauge)。由此公式得到的数对于在 10 号和 40 号之间的导线来说，误差是在 2%以内，而这两个号之间的线径覆盖了绝大多数常用的导线。需要注意的是线圈规格每增加 20，其直径就以 10 倍的比例减小，因此导线线径规格对导线直径的相对关系是与当电压用分贝表示时类似的。

把导线规格看做是直径(再强调一次，单位是英寸)的函数时，解上式得到：

$$\text{AWG}\approx 6.4-20\log d \tag{32}$$

对于漆包线来说这里没有对应的简单公式，但是可以在上面的计算结果上增加大概 0.0045 英寸来计算导线的直径，典型误差是在 5%以内或更好。需要提及的是各厂家的绝缘层厚度都不一样，因此这些公式计算得到的值，如果在严格使用的场合，必须经过验证。这里提供的这些公式主要用来指导粗略的计算。

4. 螺旋电感设计实例

下面通过优化设计一个工作于 AM 收音机频段(就是说，低于 1.6MHz)的螺旋电感，以说明如何使用以上的公式。

(1) 假设目标电感的电感值为 200μH，这是基于最终要与这个电感谐振的电容值。

(2) 选择 $l/D=5$，如之前建议的。

(3) 我们希望自谐振频率至少为 3～5MHz。这样选择的较高的值是为了增加设计余量，我们发现最大允许的自寄生电容大约是 5pF，由式(23)可以解出 D：

$$D_{\max} \approx \frac{C_{\mathrm{coil}}}{\pi\varepsilon_0\left[0.4(l/D+1)+0.9\sqrt{D/l}\right]} \tag{33}$$

这里，我们计算出最大的线圈直径为 6.4cm(大约是 2.5 英寸)，得到螺线管的长度为大约 32cm。在计算中我们忽略了引线的电容，这种考虑是有关联的，因为此例中计算得到的电感的尺寸较大，引线电容在总电容中可能会占有较大的比例。

(4) 所需的圈数为

$$n = \frac{\sqrt{L(0.9r+l)/(\mu_0\pi)}}{r} \tag{34}$$

在此例中为 131 圈。

(5) 因为对于选择的 l/D 值，最优的 d/s 值大约为 0.85，从式(30)算得最优的导线直径为 1.1mm，而线圈之间的间距为 1.3mm。

(6) 由式(32)可以计算出所需的线规为 33.5，实际上选择的线规从 32～36 都不会带来太大的偏差。

(7) 在低损耗的芯上用选好的导线均匀缠绕 131 圈，以得到具有设计好的直径和长度的螺线管。

由式(25)计算的 Q 值为 530，注意，这个计算忽略了负载损耗或辐射损耗，以及线圈绝缘层或螺线管芯的介质材料损耗。在邻近导体中感应出来的涡流损耗也未被计入。因此，这个 Q 值最好被认为是我们希望达到的最大值，而不是实际上能经常得到值。但是，上述步骤仍不失为一个电感最优化设计的方法。

5. 磁性芯上的线圈

螺线管结构的电感有一个不能自我屏蔽的问题，因此在电感和其他电路部分之间会出现一些不希望的、制造麻烦的耦合，其后果即是电路性能的降低。因而通常采用圆柱形的屏蔽罩。但是，这样的屏蔽不幸的是类似于(实际上，就是)一个次级是单圈短路的变压器。为了避免由诱生镜像电流(涡流)导致的电感值和 Q 值的严重下降，屏蔽罩的直径至少要比线圈大两倍(或更大)，以使得镜像电流尽可能远离螺线管，以致它们的效应能忽略。

另一种方式可以采用环形结构的螺线管电感。如果芯材料的导磁率(permeability)足够高，这种结构在磁性上(而不是在静电上)是自屏蔽的。磁通量基本上在芯内部，几乎不向外面溢出。但是，所有已知的磁芯材料在高频下损耗都相当大。所以，环形螺线管结构的电感只在低频下被广泛使用(一般大大低于几百个 MHz)。

大多数环形螺线管制造商都标出芯 A_L 值，这个值的单位是 mH/千圈。遗憾的是，这个约定暗指了电感值和圈数之间呈线性关系，这对不熟悉的用户是一个误导。更严格的单位应是 nH/圈数平方，而所有的数值则与 A_L 相同。

为了使环形螺线管的自谐振频率最大化，要将线圈均匀地缠在磁芯的整个周长上，但是电感两个端点之间要留下足够的距离。

许多 RF 变压器由相同圈数(或者圈数比为简单整数比)的耦合线圈组成。这些线圈是用多股导线绕成的。也即，多股相互之间是绝缘的导线作为一根单一线绕制而成。这种方

法可以最小化绕组间在寄生参数与互耦系数上的差异。

6. 单圈电感

另一个有用的公式是关于单圈电感。尽管这种结构简单，但并没有一个严格的电感值解析式表示存在（在计算总的磁通量时，会出现一些椭圆函数）。然而，一个有用的近似式子可写为

$$L \approx \mu_0 \pi r \tag{35}$$

这个式子表明，半径 1mm 的单圈电感值大约为 4nH。

在推导这个近似式子时，线圈中心的通量密度被有点任意地假设成线圈平面平均通量密度的一半，则感值可认为是总通量与电流的比值。用这个相当粗糙的近似得到的表达式工作得却不错。注意，对于圈数为 1、长度趋近于 0 的电感器，由 Wheeler 公式（式(22)）计算出的电感值收敛至 $\mu_0 \pi r$ 的 10%以内，这又独立地证明了式(35)的正确性。

一个精度好得多的表达式由下式给出，它考虑了导线的直径以及导线无限小线段之间的磁耦合：[17]

$$L \approx \mu_0 r[\ln(16r/d) - 2] \tag{36}$$

式中 d 是导线的直径。从这个式子可以看出，式(35)只在 r/d 的值大约为 10 左右才严格成立。

通过进一步将上面的粗略近似扩大化，式(35)可以被推广运用到非圆形的电感情形。我们认为所有具有相同面积的电感，不管其形状如何，都有大致相同的电感值，式(35)可改写为

$$L = \mu_0 \sqrt{\pi A} \tag{37}$$

式中，A 是线圈的面积。根据这个式子，覆盖面积为 1cm^2 的电感，其电感值约为 7nH。这个式子尽管非常近似，在估算不同形状的电感和版图寄生时被证明是很有用，同时它可以用来估测改变版图后产生的效果。

我们可以通过考虑一个半径非常大的圆周的电感，来检验以上这些表达式的合理性。因为我们能够认为这个圆圈的每一足够短的线段为直线段，可以用计算单线圈电感的公式来计算一段直导线的电感。

半径 1mm 的单圈电感的感值已经被算出来为 4nH，因此每 6.3mm（周长）长，其电感值为 4nH，这与用更精确的式子算出的结果处在相同的范围。

6.6 磁耦合的导体

6.6.1 变压器

在过去，电气工程专业的研究生至少都会熟悉理想变压器的特性。但近年来的课堂教学经验表明，现在许多学校都略去了有关变压器的内容，因此这里是个机会来补上这一课程设置的空缺（不需要复习这部分内容的读者可以跳过这一节）。我们将首先建立一个理想变

---

[17] Ramo, Whinnery, and Van Duzer, *Fields and Waves in Modern Radio*（现代无线电中的场与波），Wiley, New York, 1965, p. 311.

压器的模型，然后对它进行修改以对实际的变压器建模。

一个通常的变压器是一个磁耦合的电感器系统，变压器的名字来源于它们能够在相当广的频率范围上变换电压、电流和阻抗值。在最简单的情形下，只有两个电感——一个初级电感和一个次级电感。就像一个孤立的电感其端电压是由一个变化的磁通量引起的那样，由变压器初级产生的变化的磁通量可以在它的次级感应出一个电压；反之亦然。

对于如图6.13所示的1：n的理想变压器，n是次级与初级的匝数比。对两个电感的共同的磁通量的变化因此在次级产生为初级n倍的电压(图中表示两个极性的圆点表明哪些端点是同相位的)。能量守恒原理告诉我们，对于这一电压的升高，电流将相应地减少完全相同的倍数。因为电压对电流的比从初级到次级变化了n^2倍，阻抗的变换也以同样的倍数发生。例如，匝数比为3相应于阻抗变化为9。这个不容易实现的理想变压器能够在一个无限宽的频率范围上(包括直流)实现这一功能却没有任何损耗。尽管这样的元件实际上是不可能实现的，但正如我们很快就会看到的那样，它对于建立针对实际的变压器的模型却是一个非常有益的起点。

在前面的理想例子中，我们已经隐含地假设了由初级绕组产生的磁通量全部都耦合到次级绕组上。对大多数(但不是所有)变压器设计而言，达到尽可能接近这一理想情况是个目标。但是正如世上其他每一件事情一样，这个目标在实际中是不可能完全实现的，所以我们的模型必须接受不存在完整耦合的情况，或者说必须能容许耦合系数不为1的值。

令L_1为初级绕组自身的电感(也就是次级绕组开路时)，而L_2为次级绕组自身的电感。由这一电路的物理形态可以预期到，在任何端口上的电压应当是自感和互感项的叠加。因此对于一个(依然是)无损但非全耦合的变压器的V-I方程可以表示为

$$v_1 = L_1 \frac{\mathrm{d}i_1}{\mathrm{d}t} + M \frac{\mathrm{d}i_2}{\mathrm{d}t} \tag{38}$$

$$v_2 = M \frac{\mathrm{d}i_1}{\mathrm{d}t} + L_2 \frac{\mathrm{d}i_2}{\mathrm{d}t} \tag{39}$$

这里M为两个绕组之间的互感，它使我们能够模拟初级和次级之间的耦合程度。互易原理(这是另一个这几年来越来越不强调的概念)告诉我们，可以在初级和次级的电压公式中采用相同的M值，这对于不对称的变压器也是如此。根据电路的实际形态，互感可以取正值也可以取负值，这与孤立的无源电感不同。如果耦合过来的磁通量与自感磁通量相加，那么互感就是正值；如果它与自感磁通量方向相反，那么互感就是负值。

虽然次级或者初级两端的总电压是由初级和次级共同贡献的叠加的结果，但在以上方程中的每一项与一个平常的电感并无不同。因此，一个变压器相应的电路模型(图6.14)只含有电感元件。图中隐含地假设在两个端口之间有一个公共的连接。

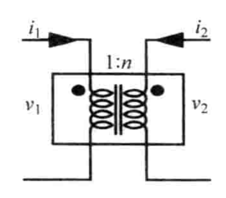


图6.13　理想1：n变压器

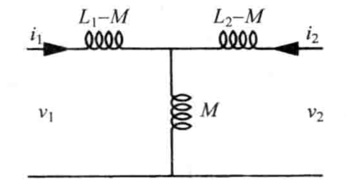


图6.14　第一轮的无损耗变压器模型(T形)

如果初级或次级相互之间非常接近，那么从一个电感产生的磁通量将几乎全部耦合到另一个电感；如果离得很远（或者如果它们的磁场方向呈正交），那么它们之间的耦合就可以忽略不计，因而 M 值很小。为此，用对耦合的定量度量来描述耦合可能发生的连续变化是很有用的，这种定量度量被合理地称为耦合系数，定义为

$$k \overset{\text{def}}{=} \frac{M}{\sqrt{L_1 L_2}} \tag{40}$$

因此，耦合系数是互感和两个个别自感的几何平均值的比。对于无源元件，耦合系数的绝对值不会超过 1。

我们的变压器的第一轮的模型看起来很不错，但也存在着一些缺点，这常常促使人们去建立其他模型。图 6.14 所示的模型的一个突出局限性是没有显式地将初级和次级绕组间匝数比包含在内，这一信息是埋在了各个电感参数内。另一个不太重要的缺陷是初级和次级共享一个公共端。但这个缺点很容易克服，只要把这个模型和一个理想的 1∶1 变压器串联在一起就可以解决，如图 6.15 所示。

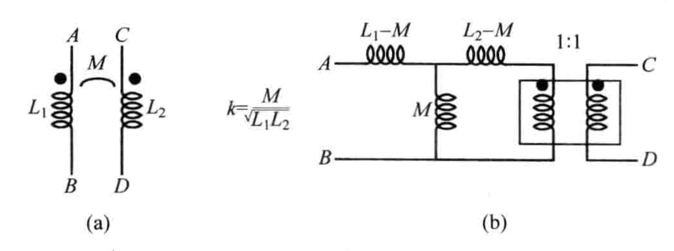


图 6.15　没有公共端的变压器电路模型

另一个既可以完全分开两个端口又可以明确包含随意的匝数比信息的模型见图 6.16，图中：

$$L_{pe} = L_1(1 - k^2) \tag{41}$$

$$L_{pm} = k^2 L_1 \tag{42}$$

以及

$$n = L_2 / M \tag{43}$$

图 6.16　另一种无损变压器模型

这个模型包含一个理想变压器作为其核心，然后用一个孤立（即不相耦合）的漏感（leakage inductance）电感 L_{pe} 来考虑不参与初级-次级耦合的磁通量。磁激励电感（magnetizing inductance）L_{pm} 则模拟了参与耦合的初级电感部分，因此它等于总的初级电感减去漏电感。这个磁激励电感还正确地说明了实际变压器是不能工作在直流（DC）情况下的，并解释了为什么低频变压器一般要比高频笨重的原因。

在耦合系数接近 1 的情况中，磁激励电感的数值一般非常接近初级电感。为了快速计算紧凑耦合的变压器电路，它们在大多数情形中都被看成是相等的。

在建立了非完全耦合和任意匝数比的无损模型后，现在需要考虑总是存在的各种寄生参数。重要寄生参数的一个可能来源是电感周围或电感在其上缠绕的材料。虽然集成电路变压器几乎从来不采用磁芯材料（所以变压器的特性基本上如同它们是缠绕在空气芯的材料上），对于分立元件电路而言，不同的芯材料则是十分普遍的。所有的磁芯材料都至少表现为有两种损耗，即磁滞损耗和涡流损耗。

磁滞损耗(hysteresis loss)来源于磁畴壁的非弹性性。为了支持磁状态的改变,这些磁畴壁必须移动。我们可以想象有一种伴随这种移动而存在的阻碍这一壁面移动的摩擦力。对于一个固定幅值的激励,每次磁状态改变时的能量损失通常可以很好地处理成一个常数,所以因这一机理所消耗的总功率近似地正比于激励的频率。可以把一个与频率有关的电阻并联在模型的初级绕组上来计入这种损耗。

涡流损耗(eddy current loss)对变压器的影响恰如对普通的导体一样。电流可以在任何邻近的导体中感应而生,包括有电导性的磁芯材料、邻近的绕组以及导电衬底等。由于感生的电压正比于频率,所以涡流损耗正比于频率的平方,这与我们在电感中所看到的情况一样。于是除绕组电阻引起的损耗外又加上了芯体损耗,而绕组电阻本身还应考虑趋肤效应。

除了考虑损耗以外,静电能还同时存储在绕组周围和绕组的电场中。因此一个高频模型还必须同时包括电容用以考虑这一附加的能量存储机理。更复杂的情况发生在当我们试图模拟高频特性且变压器的尺寸相对于这一高频波长而言不是太小的时候。在这些情形中,只对变压器进行简单集总参数电路的描述是不够的,我们必须把绕组处理成耦合的传输线。

最后,使事情变得更为复杂的是:所有的芯体材料在足够高的磁通密度下将出现显著的非线性,而且所有的参数一般都与温度有关。这些因素解释了所用的材料的多样性:没有单一一种芯体材料能满足所有的需求。

其他非理想因素最重要的影响是各种寄生参数既限制了频率响应又影响了效率。磁激励电感在直流(DC)时将使理想变压器的初级短路,使变压器不能工作,而绕组的电容在高频时也会产生一样的问题(使变压器停止工作,译者注)。这些非理想因素的净结果使经典的变压器很少能工作在千兆赫兹频率下。

6.6.2 耦合的键合线

并不是所有的变压器都是有意识地被实现的。我们需要知道围绕导体的磁场随着距离的增加下降是相当缓慢的。结果是,在相邻的(甚至是相距更为遥远的)导体间会有相当强的磁场耦合。就如在有意设计的变压器情况下,这种耦合的度量是它们之间的互感。对于两条等长度的平行而又无穷细的圆导线,这个互感可以近似地表示为

$$M \approx \frac{\mu_0 l}{2\pi}\left[\ln\left(\frac{2l}{D}\right)-1+\frac{D}{l}\right] \tag{44}$$

其中 l 是导线的长度,而 D 是它们之间的距离[18]。对于长度为 10mm、间距为 1mm 的一对导线,由上式得出的互感值约为 4nH。因为当每根导线孤立时的电感约为 10nH,所以 4nH 的互感代表互感系数为 40%。在实际情况中,一个典型集成电路封装邻近的引脚间的互感系数大致在这个数量级。此外,M 与距离的对数依赖关系意味着耦合随距离的加大而减小得相当慢,因此在非邻近管脚之间不希望的耦合也会造成麻烦。显然在高频情况下,管脚间不希望的串扰会降低信号的完整性,而且会引起一系列的麻烦。

6.6.3 宽带传输线变压器

尽管这本书集中考虑的是平面电路,但是在许多实际的场合下,一个平面电路必须在某

---

[18] 这个公式是从 Frederick Terman, *Radio Engineers' Handbook*(无线电工程师手册),McGraw-Hill, New York, 1943, Chap. 2(第 2 章)得来的。

些地方使用一个明显的非平面变压器。功率放大器就是一个例子，在那里，线路板上的变压器可能达不到要求。其他的例子包括宽带阻抗变换器，要求有超过几个频率数量级的工作带宽。使用普通的 PC 板材料的平面结构不可能提供所要求的性能指标。

我们已经研究了通常的变压器，解释了由绕组电容和泄漏电感而引起的有限的带宽。现在简略地讨论一下传输线变压器，它们（由于它们独特的结构）很少受到上面这些限制，因而能在超非寻常的大带宽下工作。

这类变压器首次是由 Guanella（1944 年）在一个几乎不可能找到的文献中发现的[19]。由 Ruthroff 给出的另一类相关的、但不同的变压器的描述则是非常频繁地被引用，这是因为它更容易被找到[20]。Guanella 和 Ruthroff 的变压器与经典的变压器的差别都是在于它们有一条从输入到输出的直流（DC）路径，而且都不在变压器的磁芯材料引入净磁通量。后面这个性质大大地减小了（依赖于频率的）磁滞损耗。而正是这个磁滞损耗设置了许多实际变压器的频率工作上限。

基本的 Guanella 结构的连接方式如图 6.17 所示。在文献中它经常被叫做非平衡变压器，但是由于变压器本身不能设定输出的共模电压（就交流而言），所以它不完全是一个典型的非平衡变压器。与输出的如何连接决定了这个共模电压。在图 6.17 所示的电路图中，将输出负载电阻的中心端接地将会使整个电路成为一个非平衡变压器。

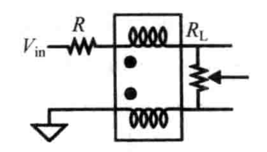


图 6.17　Guanella 的 1∶1 “非平衡变压器”（balun）

宽带运作是通过利用组成初级和次级绕组绕线的传输线本质而得以实现的。的确，Guanella 对这个结构的工作原理的解释是明确指明了对绕线作为传输线的性质的利用。因此，限制传统变压器带宽的绕组电容被吸收进变压器的基本操作。如果负载与传输线的特征阻抗相等，带宽将会被最大化。工作频率的上限因此为两个一起工作的机理而得到扩展：电抗性寄生参数被传输线常数所吸收，以及通过保持小磁芯磁通量来减少磁滞损耗。

后面这个性质可以通过考虑当图 6.17 中的上半部（初级）绕组为一个输入电压所驱动而发生的下会出现的现象，而被看清楚起了是正面的作用。对输入电压的响应而引起的输入电流的诱使在次级绕组产生电压。由于是 1∶1 的匝数比，这些绕组的电压是相同的，因而电流也是如此。但是初级和次级上流动的电流方向是相反的，因此在磁芯中的净磁通量为所企求的零，从而使基于磁滞的磁芯损耗相应地也小。的确，对于许多 Guanella 变压器的实际应用，介电损耗作用超过其他的磁芯损耗源。通常设计考虑是选择低磁滞和低传导损耗的变压器磁芯材料，但是磁芯中的介电损耗对于传输线变压器是十分重要的。

这个电路在一个很宽的频带内抑制共模信号的能力经常在仪器中被利用。当这样被使用时，这个电路通常被称作一个共模扼阻（choke）电路（见图 6.18）。它很有用，因为变压器提供的共模抑制大大缓解了对后面电路提供共模抑制的要求。例如，示波器的前端垂直放

---

[19] G. Guanella, “Novel Matching Systems for High Frequencies”（适用于高频的新颖匹配系统），*Brown-Boverie Review*, v. 31, September 1944, pp. 327-329.

[20] C. L. Ruthroff, “Some Broad-band Transformers”（一些宽带变压器），*Proc. IRE*, v. 47, August 1959, pp. 1337-1342.

大器，通常在输入连接件端和放大器的前端间插入共模扼阻电路模块。

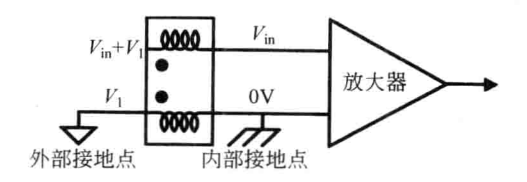


图 6.18　Guanella 1∶1 非平衡变压器用作共模扼阻电路(为清楚起见，没有画出电阻)

设想一下一个接地回路引起内部与外部接地点存在不同的交流(AC)电势。为了把问题说清楚，这里假设内部的接地点电势为 0V，并让外部接地点电压为 V_1。这两个节点之间的电压差是直接出现下面绕组的两端，使上面的绕组诱生出相同的电压。仔细考虑一下极性，从图中可以看到该诱生电压是从输入电压中减去，以使得放大器输入的就是 V_{in}。多亏了共模扼阻，地线的噪声被移除了。

在这些类型的变压器中，采用磁芯材料的目的仅仅是为了提供充足的共模电抗(因为这个电路的目标是为了纯差分输出)。成功的共模抑制的频率下限因此是为共模电感所设定。直到相应的共模电抗不再比传输线绕组的特征阻抗大，这个非平衡变压器才停止恰当的工作。

这个 1∶1 的结构通过适当组合绕组的串联和并联连接就能够被扩展到其他的阻抗传输比。图 6.19 给出了一个例子，考虑驱动两个并联的 1∶1 匝数变压器，同时将它们的输出串联连接，就可以构造出一个 1∶4 匝数比的不平衡变压器。

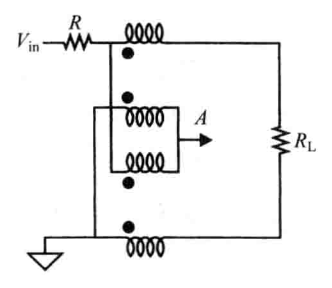


图 6.19　Guanella 1∶4 非平衡变压器

典型的电路接法是将 A 点接地。另一种接法是，可以让输出负载的中点接地(就如在 1∶1 的情况下)，这样可以使共模输出电压为零。假若变压器是用同轴线来绕制的，那么这些导体应该按照电缆中心、电缆屏蔽，再是第二条电缆的中心和屏蔽来安排，在图 6.19 的从上到下的次序。一个替代的方式是，传输线绕组可以用简单多绞线(在该情况下，是四绞线)来实现。

这个用驱动并联绕组及串联的绕组上得到输出的一般想法可以很容易地被推广来实现 $1:n^2$ 的平非衡变压器，其中 n 为整数，等于被使用的 1∶1 结构的基本模块。

而 Ruthroff 结构有些不同，就如我们会看到由于这儿的带宽较差，我们将只考虑基本的 Ruthroff 的 1∶4 变压器的两个例子：一个是非平衡变压器，另一个则不是。

在图 6.20(a)中，其结构有一个非平衡的输入和输出，因此有时称之为“unun”。但是不论你多少次大声说(而且不管你的母语是什么)听起来都很别扭。它写出来也是看起来傻傻的。不管怎么说，它的工作原理如下，一个外加电压加在下面的这个绕组则在上面那个绕组感生出同样大小的电压。这两个绕组因为是串联的，从而使输出电压加倍，进而产生 1∶4 的阻抗传输比。这个结构有时也称为自举连接，因为串联的接法帮助了输出电压的增加。

图 6.20(b)的结构是将电压加在上面的线圈上，从而在下面的线圈上感生出同样大小的电压。再一次，我们跟踪极性，看到电阻负载的左边(即上端，译者注)见到一个等于输入

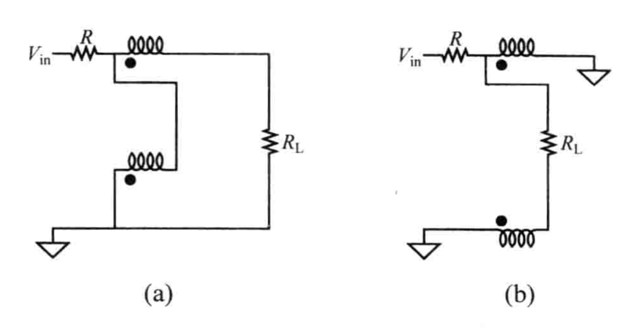


图 6.20　Ruthroff 1∶4“unun”(非平衡输入、非平衡输出)(a)与非平衡变压器

电压的电压，而电阻的右端(即下端，译者注)见到的是这个电压的负值。结果是负载上的电压又是输入电压的两倍，产生了一个 1∶4 的变换。因为负载电阻的两端都不接地，这个结构被看作是一个平衡(输出)结构。

Ruthroff 版本有更为局限的带宽，这是因为这些结构将一个外加的电压与一个沿着一根(传输线)绕组传输的电压相加。后者必然会有延迟，因而产生了一个基本的频率上限。这种变压器只有在延时短于信号的周期时才能正常工作。将绕组线缩短可以改进频率的上限，但是因为减少了共模电感而同时使频率的下限变差。采用磁导率高的磁芯可以在某种程度上得到补偿，但是没有多少磁芯材料既具有低介电损耗又有高磁导率。

在 Guanella 的结构中，输出是从两根传输线的两端得到的。因此延时是共模的，不会对频率的上限发生影响。要了解更多的关于该类型变压器的信息，可以参见 Jerry Sevick 的 Transmission Line Transformers(传输线变压器)，2nd ed.，American Radio Relay League，Newington，CT，1990.

6.6.4　窄带传输线变压器

如果不要求宽带工作，阻抗变换可以用更加简单的传输线变压器来实现。例如，一个应用十分广泛的 1∶4 窄带非平衡变压器可以容易地用一个半波长传输线来构造。

在图 6.21 中，一个阻值为 R 的电阻器与半波长传输线在两端相连接。为了分析简单，我们将整个电阻分成两个电阻串联连接。由于传输线的整个特别的长度，加在传输线左端的电压 V，当传输到传输线右端时，经历了一个反相。电阻 R 的中点因此处于地电位。传输线的右端的负载电阻也因此为 $R/2$，这个负载电阻反射到传输线的左端时依然是 $R/2$。这个被反射回去的负载电阻与另一个已经在传输线左端的 $R/2$ 电阻相并联，结果是产生了一个净输入电阻 $R/4$。因此，一个平衡的负载电阻 R 被转换下降为一个非平衡输入电阻 $R/4$(反之亦然)。

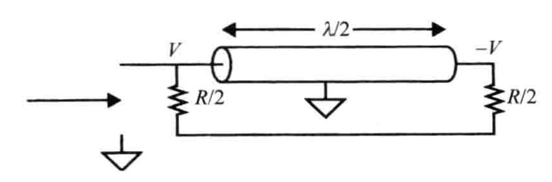


图 6.21　窄带 1∶4 传输线非平衡变压器

为了便于平衡负载的连接，通常是将传输线弯成 U 形或 O 形。这个非平衡变压器在频率高到用微带线来实现变得可能时是特别有用的。

最后一个将要考虑的不平衡变压器是一个由同轴电缆设计转换成带状线的结构，它最先是由 Nathan Marchand 提出的[21]。它可以被认为是像两个连接在一起的耦合器(couplers)但被反相激发。可以从图 6.22 看到，一个不平衡输入驱动一个终端开路的 $\lambda/2$ 传输线。因此信号在传输线两端是正好反相的。这两个反相信号驱动了一对耦合器，因此它们的输出同样也是反相的。因为这对传输线每条的末端接地，因此差分输出都是所需要的平衡输出。

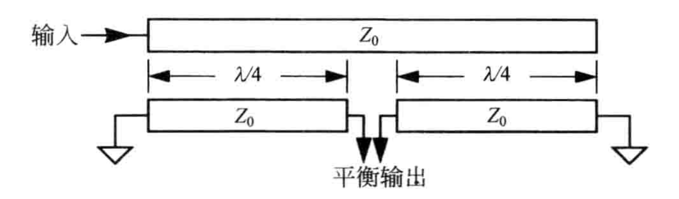


图 6.22　经典的 Marchand 非平衡变压器

诸如带状线的侧边耦合的结构并不十分适合用来实现在一个 Marchand 非平衡变压器通常希望的紧耦合。假若额外的金属层可以被利用，那么可以用宽的侧边耦合的版图来提供所需要的紧耦合，而不求助于近于荒唐的十分小的带状线的间距。

尽管经典的 Marchand 不平衡变压器是窄带结构，它依然是可以实现宽带的，这可以通过使用多段实现的传输线来代替如图 6.22 所示的单根 $\lambda/4$ 段(见第 7 章)。然后可能实现带宽超过几个两倍或者十倍。最后，通过调整两条互相耦合的传输线之间的非耦合部分的长度，能够实现输出相位平衡的微调。

6.7　总结

我们已经看到，随着频率的升高看起来十分普通的元件必须被复杂度不断增加的方式来进行建模。通常是简单的元件有一些重要的行为只能在低频下才能被忽略。对于微波电路的恰当设计，即使是电阻、电容和电感都必须作为复杂的阻抗来对待。这一章提供了大量的公式和估算寄生电感与寄生电容的规则，以建立合适的模型。最后，我们还考虑了各种类型的变压器，尽管这些元件很少是用纯平面结构实现的。

[21] N. Marchand, "Transmission Line Conversion Transformers"(传输线转换变压器), *Electronics*, v. 17, December 1944, pp. 142-145.

第7章 微带线、带状线与平面无源元件

7.1 引言

在射频设计中，一个不断出现的课题是需要仔细考虑沿着信号通路的每一个元器件的电气特性。这种考虑也延伸到印刷电路板，所以本章研究印刷电路板的高频特性以及由印刷电路板材料制成的众多的无源元件的高频特性。我们将特别关注一种被称为微带线的专门的传输线类型，微带线尤其适用于平面微波电路的实现。此外，传输线还可以形成许多种无源元件，所以我们也讨论电容、电感、谐振器、功率合成器及各种耦合器——包括巴伦(baluns)，即非平衡变压器和混合型器件(hybrids)。

7.2 印刷电路板(PC 板)的一般特征

就像应用在低频领域的印刷电路板(PCB 板)那样，应用于射频的 PCB 板也是由某种绝缘层隔开的一些金属层构成的。按照美国的习惯，金属层厚度是由每平方英尺的铜的重量间接给出的。这样，"1 盎司"铜(一个常用的值)大约是 1.34mil(35μm)厚[①]。0.5 盎司和 2 盎司铜也是常用值。铜块的直流电阻率大约为 1.8mΩ·cm，因此对应的 1 盎司铜的方块电阻近似为 0.5mΩ/□。1GHz 时铜的趋肤深度是 2.1μm，此频率下铜层的方块电阻增大到大约 8mΩ/□。根据铜层形成和淀积的方式的不同，电阻率可以比在铜块中的值大到 2 倍[②]。另一个影响电阻率的因子是表面粗糙度，如果表面粗糙，电流就得流过更大的距离。对这种电阻升高的倍数的估计值为[③]：

$$F_{sr} \approx 1 + \frac{2}{\pi}\arctan\left[1.4\left(\frac{\Delta}{\delta}\right)\right] \tag{1}$$

其中，Δ 是粗糙度的均方根值，δ 是趋肤深度。从这个表达式可以看到当均方根值与趋肤深度可以比拟时，表面粗糙度就变为一个十分显著的因素。

通常的电介质层厚度是 1/32 英寸(0.8mm)，1/16 英寸(1.6mm)，也有时是 1/8 英寸(3.2mm)。在多层板中，也会遇到 1/64 英寸(0.4mm)的情况。这些都是近似值，在美国实际厚度通常标为 5 或 10mil 的倍数(在用公制的国家中则取相应的整数)。

对于低频领域的一般应用，最常用的 PCB 板的电介质材料是被称为 FR4(flame

---

① 1mil 是 0.001 英寸，不要与 1 毫米混淆；1mil 实际上大约是 25μm。

② 由于趋肤效应使面向接地导电层的导体表面的电流密度为最大，这个表面的粗糙度与电阻的计算最为有关。通常的情况是金属导体层是经由助黏剂粘到基板上的。这个助黏剂和金属的接触面会是相对粗糙的，这样就会在高频时产生令人惊讶的高导电损耗。

③ E. O. Hammerstad and F. Bekkadal, *A Microstrip Handbook*(微带线手册), ELAB Report, STF 44 A74169, N7034, University of Trondheim, Norway(挪威), 1975.

retardant formulation number 4，阻燃配方编号4)的玻璃纤维环氧树脂，或者它的较易燃的对应材料G10。FR4的介电损耗在微波频段中比导体损耗更为严重。对于1/16英寸(1.6mm)的典型的微带传输线(这将在7.3节详细讨论)材料，损耗是0.03dB/(cm·GHz)。由于介电损耗是主要的损耗机理，在dB尺度上，衰减在一个很宽的频带内(从大约50MHz到至少5GHz)随频率线性增加。这种函数依赖关系说明了在任意的频率下，单位波长的损耗是常数。在现在这种情况下，这个常数是0.5dB/波长。

除了这个有点高的损耗，FR4板制造的精度不高，所以，一般认为FR4不适合性能要求高的射频电路的大量生产。例如，过去几年内斯坦福大学从多家供应商购买的1.6mm厚FR4板，其相对介电常数在1GHz时分布在4.2～4.7之间。这种分布不是均匀的，但如果认为这一不是十分科学的、稀疏的采样还是具有代表性的，我们可以推断介电常数值在标称值4.5左右的变化超过5%[④]。尽管统计理论是怎么说，通常人们需要预计到更大的偏离标称值的范围，比如说±10%。如果这种变化范围是允许的，而且损耗也能被容忍，即便是传统的看法认为FR4不适合应用在几个GHz以上，FR4仍可以用于5GHz以上的频率范围(一些胆大的人甚至已经用在10GHz上)。

应当注意到平面传输线是浸入在非均匀的电介质中(例如，部分是FR4，部分是空气)。采用介电常数的加权平均就可以计入这个效应，而且使表达式简单，这至少看起来是如此。然而在高频时，具有更高(比空气，译者注)介电常数的衬底会吸引更多的电通量到衬底部分，从而使有效介电常数升高。这样高频信号就比低频信号传播的慢，因此这种传输线有色散的性质。当频率从100MHz升至5GHz时，FR4的有效介电常数可以预计提高大约5%的一个典型值。这种色散引起的时域波形失真可能会限制非常注重脉冲保真度的宽带应用(例如高速数字电路)[⑤]。

由于FR4是在环氧基矩阵内将玻璃纤维结合在一起构成的，如果纤维走向不是随机的，有可能会出现各向异性。例如，不同方向的介电常数不同的材料也不是前所未闻，以至于相互正交的传输线可能有不同的特征阻抗。虽然FR4有各种各样的缺点，但它极低的成本与到处有供应的优势，不断地促使工程师想方设法在非关键、带有业余性质的应用以及小量样品的射频电路原型的应用中采用它。这里可以提到一个相关的事实，即数字电路的时钟频率已经达到了此前被称为微波的领域。

在极低成本的消费电子设备中(例如，玩具、袖珍收音机等)，更廉价的电路板材料也不罕见。酚醛通常是焦糖棕色，有典型的"有机化学物质"气味，损耗非常大。虽然酚醛有时用于频率高达100MHz的玩具，但它完全不适合要求严格的应用。在这里提及这点仅仅是为了回答这个问题："这廉价而且有恶臭的板子是什么做的?"

当然在要求高的场合，必须采用微波级别的材料。很多质量最好的软衬底材料几乎都是基于PTFE(聚四氟乙烯，杜邦公司的Teflon™品牌更为人们所熟知)。然而用这种材料制造多层板是相当困难的，为了解决这个问题，提出了一些替代方法。一个被广泛使用的例

---

④ 使事情更为复杂的是，对于非常薄的FR4衬底通常使用一种不同的环氧树脂与玻璃纤维的比例，这使得这种用于层压结构的材料的介电常数还要低5%～10%。

⑤ 这种特性没有妨碍不受拘束的工程师使用均衡器来降低色散的影响。通过这种方法，FR4线也可传输特高带宽的信号。

子是 Rogers 公司的 RO4003(http://www.rogers-corp.com/mwu)。这种材料基于玻璃织物增强型碳氢化合物和陶瓷热固性塑料材料。基于 RO4003 的传输线的损耗大约只有以分贝计算的 FR4 损耗的 1/4。其介电常数为 3.38，误差可控制在 1.5%以内。低损耗、稳定、窄的变化分布特性以及容易制造等优点，使这种材料特别适合许多各种应用。

有时，也可能遇到聚苯醚(polyphenylene oxide，PPO)材料的印刷电路板。例如，PPO 的较低的损耗(相对于 FR4)特性使其能够用于一些高频 Tektronix(美国一家电子仪器公司，译者注)示波器的插件。如果你真的遇到这种材料，一定要小心对待，因为 PPO 的熔点很低，这就像高科技的黄油。手工焊接或去焊组件仅勉强可以做，但必须以你能掌握的尽可能快的速度来完成。同时要小心，在焊接时不要加力，否则会产生物理形变。

本文以后主要关注 FR4(也有小部分是有关 RO4003)，所以，在表 7.1[⑥] 中我们总结了它们的一些相关特性。关于 FR4 的有些条目是猜测的，因为实际上很少有参数是可控的——这可以从 FR4 相应的标准 IPC-4101 的相当宽松的规定中看出。

表 7.1 FR 与 RO4003 在 300K 下的一些特性

性　质	FR4	RO4003
1GHz 下的体相对介电常数 ε_r	典型值：4.5(1±10%)(1MHz 下，最大值 5.4，根据 IPC-4101 标准)	3.38(1±1.5%)
ε_r 的温度系数 $T_C/(10^{-6}/℃)$	?	+40
损耗因子	典型值：1GHz 下 0.015(1MHz 下，最大值 0.035，根据 IPC-4101 标准)	10GHz 下 0.002
尺寸的稳定度	$<500\times10^{-6}$	$<300\times10^{-6}$
热膨胀的温度系数 $T_C/(10^{-6}/℃)$	典型值：在板的平面方向：15，垂直(z)方向：100	在板的平面方向：15，垂直(z)方向：50

关于 FR4 的最后一条注释是，知道它是一种极好的绝缘体有时也是很有帮助的。除了表现出较低的表面泄漏(若表面清洁则表面电阻的典型值是 200GΩ)，它的击穿电压也很高。1.6mm 厚的衬底的典型承受能力可超过 40kV(有些会达到 60kV 以上)。很好的耐高压电容可以用 FR4 手工制造而成。

在要求更为苛刻的应用中，可以使用各种硬质衬底，如氧化铝、氧化铍、石英和蓝宝石。硬质衬底具有比我们讨论过的软质衬底更好的热导性能，但前者更为昂贵。氧化铝是最常用的硬质材料，因为它是上述硬质衬底材料中最便宜的，而且电学特性也还不错。金属能很好地粘附于其上，所以形成互连不成问题。机械加工也较为容易。同时它也非常坚硬，可以附着在其他热系数不相匹配的衬底上。这种硬质衬底也可以抛光到十分平整的表面，使得通过金属淀积而形成的互连具有相似的低的表面粗糙度，从而降低高频导电损耗。氧化铝的介电常数(大约是 10)兼有利弊。低频时可以形成比 FR4 更为紧凑的电路，但在高频时却迫使采用不容易实现或根本无法可靠地制作的电路尺寸。就如我们将进一步简要地讨论指出的：更高的介电常数意味着在导体中传播的电磁波在导体表面会经受到更大的非均匀性，这对高频性能有负面的影响。

---

⑥ 另外一些电介质材料及其相应的介电常数见附录 C(7.10 节)的表。

氧化铝，蓝宝石的晶体型有利于抛光得更细腻。所以，当不考虑成本并且单纯追求降低表面粗糙度到最小值时，会偶尔使用它。但矛盾的是，极度的光滑却有碍于金属的附着，因而在蓝宝石表面和实际互连金属之间必须加入助黏层。如果粘附层本身不平滑，即可使蓝宝石的光滑优势丧失。

由于往往希望能减小源自介质不均匀性的问题，有时也会用熔融石英替代氧化铝或蓝宝石。其介电常数为3.8，使微波（或更高频率的）电路所需的加工公差相当宽松。其介电损耗在目前使用的硬质衬底材料中也是最低的，在微波波段也能保持很低。遗憾的是，石英非常易碎，因而无法进行机械加工。将其制作成衬底表面时必须极其小心，以免热膨胀系数不同使其破碎。石英非常光滑，因而具有蓝宝石在这方面的全部优缺点。总之，使用这种材料难度是很大的。

最近，低温共烧陶瓷（LTCC）衬底受到欢迎。比如说，这种材料的损耗介于FR4和RO4003之间，但比传统的微波衬底，如氧化铝，更容易制造。

我们再提及一种材料：因氧化铍具有良好的热导性能和与铜非常匹配的热膨胀系数，一度广泛用于高功率模块中。但是吸入氧化铍的有毒粉尘对于某些人可能会致命，所以不要对其进行机械加工。最好是完全避免使用这些材料——但如果你不能避免，也绝对不要打磨或碰碎BeO！

7.3 印刷电路板上的传输线

印刷电路板的平面特性使其不能制成同轴结构。此外，与同轴线的中间点相连是非常困难的。1951年，Robert M. Barrett提出了用印刷电路板制作方法实现许多经典微波元件的平面版本[7]。美国国际电话电信公司（ITT）联邦电信实验室的研究人员发表了三篇论文，进一步发展了这一通用的想法（见1952年12月的IRE会刊，以及1955年3月的IRE的微波理论与技术学报，MTT）[8]。正是这些以及后续的努力才产生了被称为带状线和微带线结构等平面传输线（见图7.1和图7.2）。虽然在该实验室外普及还需要很长时间，它们现在是迄今最为常用的构造微波电路的方式。

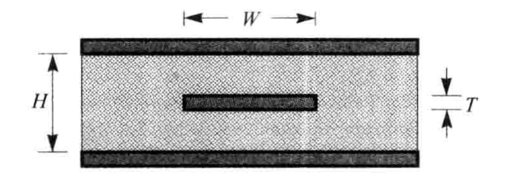


图7.1 带状线（截面图）

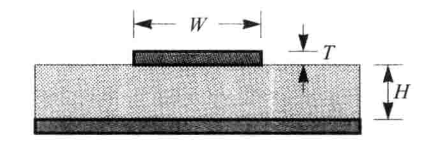


图7.2 微带线（截面图）

⑦ R. M. Barrett and M. H. Barnes, "Microwave Printed Circuits"（微波印刷电路），*Radio and TV News* (Radio-Electronic Engineering Section), v. 46, September 1951, pp. 16-31. Barrett归功于V. H. Rumsey和H. W. Jamieson在第二次世界大战期间首度在一个功率分配器使用了平面线（带状线）。也有传说是多产的Harold Wheeler在20世纪30年代后期已经进行了共面线实验。

⑧ 或许这三篇文章中最卓著的是D. D. Grieg和H. F. Engelmann的论文，"Microstrip-A New Transmission Technique for the Kilomegacycle Range"（微带线——一个用于千兆周期距离的新传输技术），*Proc. IRE*, v. 40, December 1952, pp. 1644-1650. 这篇文章标志了微带线这个词首次出现在文献中。ITT已经收购了联邦电信（其因电弧技术而著称），正是这个组（被更名为联邦电话）发表了这个有关微带线的研究工作。

如图 7.1 和图 7.2 所示，带状线是由夹在两根地线间的一根导线构成的，而微带线是由位于单个接地平面之上的导线构成。带状线基本上是自屏蔽(只在侧面有些泄漏)，因此，在需要使辐射造成的损耗和不期望的耦合尽量小的场合非常有用。此外，由于导线是置于均匀电介质中，因而支持 TEM 传播。然而，带状线的一个重要缺点是不易与中心导线形成连接。

图 7.1 所示的带状线是最为普遍被实现的结构：一根导线置于两个接地平面中间，形成一种对称结构。这种线的特征阻抗为[⑨]：

$$Z_0 \approx \frac{60}{\varepsilon_r}\ln\left[\frac{6H}{\pi W(0.8+T/W)}\right] \tag{2}$$

这个表达式对于比较窄的带状线，即 $W/(H-T)<0.35$，最为准确。对于使用 1 盎司铜，电介质总厚度为 1/8 英寸(约 3.2mm)，一根特征阻抗为 50Ω 的带状线所需的宽度约是 1.25mm。尽管这个宽度依据我们上面给出的不等式看上去不是那么窄，方程(2)仍有良好的精度，至少对于阻抗在 50Ω 附近的带状线而言是如此。

如需更精确的估计，可使用 Cohn 的最初较为复杂的表达式[⑩]：

$$Z_0 = \frac{\eta K(k)}{4\sqrt{\varepsilon_r}K'(k)} \tag{3}$$

其中，

$$k = \left[\cosh\left(\frac{\pi W}{2H}\right)\right]^{-1} \tag{4}$$

η 是自由空间的阻抗，

$$\eta = \sqrt{\mu_0/\varepsilon_0} \approx 120\pi \tag{5}$$

此外，以下的近似表达式的精度是如此之好，我们可以将其视为等式：

$$\frac{K'(k)}{K(k)} = \begin{cases} \left[\dfrac{1}{\pi}\ln\left(2\dfrac{1+\sqrt{k'}}{1-\sqrt{k'}}\right)\right], & 0 \leqslant k \leqslant 0.707 \quad (6) \\ \left[\dfrac{1}{\pi}\ln\left(2\dfrac{1+\sqrt{k}}{1-\sqrt{k}}\right)\right]^{-1}, & 0.707 \leqslant k \leqslant 1 \quad (7) \end{cases}$$

其中，$K(k)$是完全椭圆积分(第一类)，K'是它的余函数(假若你们想知道 K，K'代表什么)。最后，还有一个有用的关系式：

$$k' = \sqrt{1-k^2} \tag{8}$$

然而，比带状线更为常用的是微带线——即使它既不是一个被屏蔽的结构也不是一个均质结构。要注意传播不可能是纯 TEM 的，这是由于以下这种非均质性：导线上方材料(空气)的介电常数与下面的不同。为满足介面处的电场边界条件，需要一个沿传播方向的场分量。笼统地说，在导线下的那部分波“想要”以低于导线上那部分波的速率传播。于是这种非均质结构的理论推导非常复杂也是不足为奇了。因此，通常用“准 TEM”，作为一种

---

⑨ 这个表达式是对 Seymour B. Cohn 最初提出来的表达式的一个修正。“Characteristic Impedance of the Shielded-Strip Transmission Line”(被屏蔽的带状线为传输线的特征阻抗)，*IRE Trans. Microwave Theory and Tech.*, July 1954, pp. 52-57.

⑩ 同时可见文献 K. C. Gupta, R. Garg, and R. Chadha, *Computer-Aided Design of Microwave Circuits*(微波电路的计算机辅助设计), Artech House, Dedham, MA, 1981. Cohn 的公式是基于保角变换技术，没有考虑到非零导体厚度的情形。

简化的概念。在此概念中，结构等效为均质材料包围的一根导线，而这种材料的介电常数已经作了相应的其值变小的处理，以得到一个有效的(平均)介电常数。

虽然有些缺点，微带线是十分常用的结构，因为其导线不是埋起来的，因此易于构建连接。制作样品 PCB 板的标准方法涉及刻蚀表面覆盖着铜层的 FR4 板，其需要的走线图形则在化学光刻胶上定义。另外一种快速的方法且不用有毒性的材料是，将有黏性的铜箔胶带直接贴成所需图案。使用一种工具刻刀对铜箔进行需要尺寸的修整是最为理想的。这种制作方法显然只限于频率足够低的应用，那里尺寸的(不)稳定性是可以容忍的。

铜箔胶带在电子产品商店里通常是一个特别订购的商品，在大多数连锁电子商店里一般并没有销售。然而，从另外一个令人想不到的地方则能买到它：五金店和园艺用品店，因为，铜箔胶带可以作为阻挡蜗牛的屏障。SureFire™ Slug & Snail 是一个颇具代表性的铜隔离胶带品牌，在 Ace 五金店这种连锁店有售。在使用铜箔胶带前要用塑料垫轻轻打磨电路板，去掉表面的保护膜(金属垫也能达到同样的效果，缺点是会在电路板与工作台上留下金属碎屑)。然后在 FR4 上将铜箔胶带刻成所需的尺寸和形状。用像笔管这样的塑料物体把胶带压平整，并作必要的修整。就是这么简单。除却简单，这种方法的美还在于易于进行修改。如果你出了差错(如刻刀切下过多的材料)，没关系，揭下胶带，再次清洁电路板，再贴一条新的胶带就可以了。无论是用这种方法还是传统的印刷电路板制造方法，电路元件都可以方便地和导线连接，这有利于制造、检测、调整、故障排除和维修。

有个始料未及的挑战是关于这些导线的特征阻抗的已经发表的表达式是多种多样的。所幸的是，通过比较发现这些表达式的估计值仅相差百分之几(至少对于 50Ω 附近的阻抗是这样的，对于其他阻抗误差可能要大些)。对于微带线情况，一个代表性的公式为[11]：

$$z_0 \approx \sqrt{\frac{\mu_r}{\varepsilon_r}\frac{\mu_0}{\varepsilon_0}} \cdot \frac{H}{W} \cdot \left[1 + 1.735\varepsilon_r^{-0.0724} \cdot \left(\frac{W}{H}\right)^{-0.836}\right]^{-1} \tag{9}$$

对于几乎所有情况，磁导率 μ 是自由空间的磁导率，为 $4\pi \times 10^{-7}$ H/m，或近似为 1.257μH/m。

表 7.2　在 FR4 PCB 板($\varepsilon_r = 4.4$)上的微带线有代表性的物理宽度

H	W for 50Ω	W for 100Ω	f_c@50Ω
1/16 英寸(1.6mm)，1-oz. Cu	3.0mm	0.65mm	26GHz
1/32 英寸(0.8mm)，1-oz. Cu	1.5mm	0.30mm	48GHz

注意到式(9)忽略了导体厚度依赖关系。因此，该公式会产生一定误差，即便这些误差的量通常是完全可以忽略的。为了得到一个初级近似的表达式，我们假设导体的垂直侧面的贡献就像是它们被折叠放平而增加了宽度。这样，就可以认为有效宽度的增加量大约

---

[11] R. S Carson, *High Frequency Amplifiers*(高频放大器), 2nd ed., Wiley, New York, 1982, p. 78. 一个更精确(因而是趋于无限的复杂性)的表达式见：E. Hammerstad and O. Jensen, "Accurate Models for Microstrip Computer-Aided Design"(微带线计算机辅助设计的一些精确模型), *IEEE MTT-S Digest*, June 1980, pp. 407-409; R. H. Jansen and M. Kirschning, "Arguments and an Accurate Model for the Power-Current Formulation of Microstrip Characteristic Impedance"(微带线特征阻抗的功率-电流公式中的自变量与一个精确模型), *Archiv für Elektronik und Übertragungstechnik*, v. 37, no. 3/4, March/April 1983, pp. 108-112; 以及 M. Kirschning and R. H. Jansen, "Accurate Wide-Range Design Equations for the Frequency-Dependent Characteristic of Coupled Microstrip Lines"(耦合微带线的频率依赖特性的一些精确、宽应用范围的设计方程), *IEEE Trans. Microwave Theory and Tech.*, v. 32, no. 1, January 1984, pp. 83-90. 这些准经验方程是从大量电磁场仿真数据拟合中得到的。

是 2T。

如果要求更高些，以下方程可用于计算出更为精确的有效宽度：

$$W_{\text{eff}}=W+\frac{T}{\pi}\cdot\left[\ln\left(\frac{2H}{T}\right)+1\right] \tag{10}$$

无论是怎样算出来的，有效宽度都被用在公式(9)中。举个具体例子：一块 1/32 英寸(0.8mm)厚的一盎司铜膜电路板(如前所述其厚度约是 35 μm)。假设衬底介电常数是 4.4，这样，公式(10)对有效电气宽度的数学修正仅仅是增加约 1.5T，或者是 2.0mil(50 μm)。这种修正与典型的制造公差(如 3mil)在大小上是相似的，并且一定比用刻刀和铜箔胶带方法的公差小。因此，通常可忽略。上例中，对于 50Ω 的导线(见表 7.2)物理宽度须近似为 1.5mm，或大约 58mil(即 0.058 英寸)。在 1/16 英寸板上，恰好可以用刻刀把 1/4 英寸宽(在美国这是一个标准的宽度)的铜箔胶带从中间割开，从而得到十分接近于 50Ω 的阻抗。

如前所述，1/16 英寸厚的 FR4 材料上的 50Ω 微带线的损耗的典型值约是 0.07～0.09dB/(in・GHz)(大约 0.03dB/(cm・GHz))，而 RO4003 材料是这个值的 1/4[⑫]。图 7.3 是 50Ω 微带线传输性能随频率变化一个典型曲线。在此情况下，相应的损耗约是 0.03dB/(cm・GHz)。如图所示，在一个很宽的频率范围内(这里，频率变化超过 2 个数量级，即 50MHz～5.05GHz)，衰减确实随频率线性增大(以 dB 表示)。

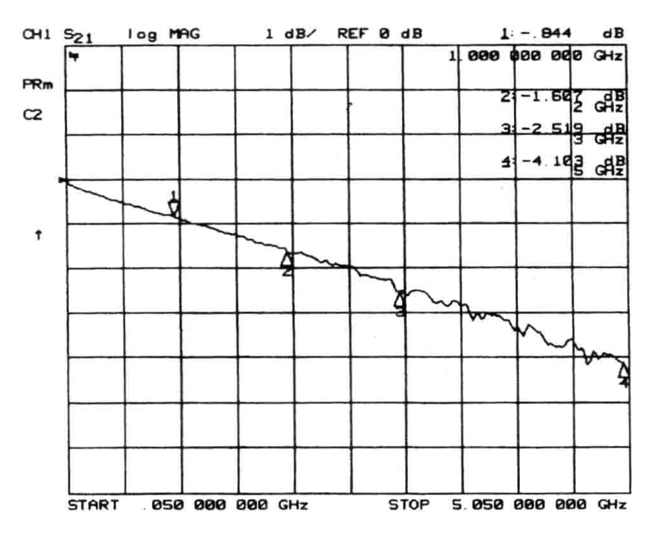


图 7.3 在 1.6mm 厚 FR4 上的 50Ω 微带线的典型传输系数(S_{21})与频率的关系

就如与连接件和其他任何可以传递电磁能量的元件，我们必须知道在传输线传播时可能产生的模式转换(moding)。对于微带线(以及带状线)，这种模式转换可出现于或者是频率足够高以至于半波长恰好等于传输线宽，或者是当 λ/4 大约等于传输线到接地平面的距离[⑬]。前一种情况描述了横向模式谐振，而后一种情况相应被传播的能量可以耦合到表面波的条件。这种表面波可沿着电介质或在电介质内传播。

由于微带线具有非均质性，很难严格推导这些界限。但可以通过忽略这个非均质性，来

⑫ 要继续记住 FR4 材料的变动性。同时要再次注意这种损耗行为意味着单位波长的损耗在 dB 表示下是常数。

⑬ 在此情况下传输线和其镜像间的距离就是半波长，所以模式转换的界限都可以用半波长表达。

给出一个非常粗略的估计。这样，第一个横向谐振产生的波长可由以下公式近似计算：

$$\lambda_c \approx 2W\sqrt{\varepsilon_r} \tag{11}$$

相应的频率是

$$f_c \approx c/2W\sqrt{\varepsilon_r} \tag{12}$$

边缘效应导致传输线在电学上的宽度要大于 W。

推导表面波模发生的公式则更费劲。然而继续用粗略规则，可以非常近似地估计这个表面波模频率为

$$f_T \approx c/4H\sqrt{\varepsilon_r} \tag{13}$$

一种更为严格的推导给出[14]

$$f_T = \frac{c}{2\pi H}\frac{1}{\sqrt{\varepsilon_r - 1}}\arctan(\varepsilon_r) \tag{14}$$

可以看出那个近似表达式(13)是有点保守的，其预测的在表面波产生的频率比实际值要低些。这两个表达式在 ε_r 趋于无穷时则趋于一致。

对于 FR4 上的 50Ω 传输线，W/H 值近似等于 2，因此对应于这两个条件的截止频率互相接近。为了提高没有高阶模式的带宽，需要有更小的尺寸，而同时为了维持传输线阻抗不变，衬底厚度因而要随线宽减小而减小。而伴随的副作用是较窄的线有更高的导体损耗，所以工作的频率越高，传输线的衰减会越大。这种不尽人意的性质对于所有传输线都是存在的(例如，微带线、带状线、同轴线)，也正是这一特点促使了在更高的频率时使用波导。

表 7.2 的最后一列给出了对应于横向谐振频率的近似值。对于低于截止频率的工作频率，比如低大约 5%～10%，良好的阻抗特性依然被保留。最后，请注意即使是 1/16 英寸 FR4 也有足够高的无高阶模的带宽，因此可以工作在该材料所适用的任何频率上。

7.3.1　共面波导(CPW)和共面带线(CPS)

虽然微带线是十分方便的，但它也有些缺点。一是有大量能量可能耦合到衬底，如果衬底带有损耗，就会导致能量消耗。另一个是量测微带线的微小结构过于困难，这是因为接地平面在电介质的另一面。而对地的连接也不容易做好，因为存在着接地通孔导体的寄生电感。如果通过在电路上打大量接地通孔导体可以解决这一问题，但同时也会增加制造复杂度和成本，所以这不是一种非常实用的解决方案。

另一种方法是使用共面导体，这样，地线和信号线都在电路板表面。不言自明，共面放置大大有利于微波及以上频率的量测。若三条导线以地线—信号线—地线方式排列，这就是共面波导(CPW)或共面传输线。若一对导线以地线—信号线方式排列，就是共面带线(CPS)，见图 7.4。

这些结构还有另一个优点：不同的导线宽度和间距都能产生相同的特征阻抗。

因此，除非需要连接尺寸大的连接件，导线都可以保持较小的尺寸。因此，CPW 提供了一个极佳的到同轴连接器的过渡。这种共面结构的性能随导线宽厚比减小而提高，这是因

[14] Vendelin, "Limitations on Stripline"(带状线的局限), *Microwave Journal*, v. 13, no. 5, 1970, pp. 63-69.

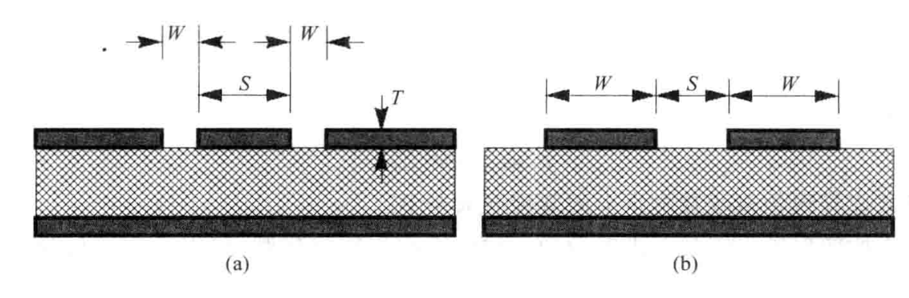


图 7.4　共面波导(a)与共面带线(b)

为边缘耦合的增加意味着耦合到带损耗的衬底的能量损耗减少。因为同样的原因，同时色散也降低了。

从图 7.4 可以看出这两种结构在某种意义上是互补的。这种互补性通过对 CPW 线的间隙宽度和 CPS 线的导线宽度都采用同样的 W，可以在形式上得以证实。就像我们将要看到的那样，求特征阻抗的这些表达式也反映了这种互补性。正如在微带线的情形下，由于在推导中使用了保角变换，在表达式中出现了椭圆函数。“保角变换”和“椭圆函数”这两个术语事实上意味着“无法推导下去”和“很难进行计算”。仔细看下面的这些方程可以消除对这个说法的质疑[15]。

$$Z_0 = \begin{cases} \dfrac{\eta K'(k)}{4\sqrt{\varepsilon_e}K(k)}, & \text{CPW} \quad (15) \\ \dfrac{\eta K(k)}{\sqrt{\varepsilon_e}K'(k)}, & \text{CPS} \quad (16) \end{cases}$$

这些看起来简单，但看到子表达式就会觉得很复杂了。首先，

$$k = \frac{S/2}{S/2+W} \tag{17}$$

$$\varepsilon_e = 1 + \frac{\varepsilon_r - 1}{2}\frac{K'(k)K(k_1)}{K(k)K'(k_1)} \tag{18}$$

继而，

$$k_1 = \sinh\left(\pi\frac{S}{4H}\right)/\sinh\left(\pi\frac{S/2+W}{2H}\right) \tag{19}$$

然后，又一次有

$$\frac{K'(k)}{K(k)} = \begin{cases} \left[\dfrac{1}{\pi}\ln\left(2\dfrac{1+\sqrt{k'}}{1-\sqrt{k'}}\right)\right], & 0 \leqslant k \leqslant 0.707 \quad (20) \\ \left[\dfrac{1}{\pi}\ln\left(2\dfrac{1+\sqrt{k}}{1-\sqrt{k}}\right)\right]^{-1}, & 0.707 \leqslant k \leqslant 1 \quad (21) \end{cases}$$

其中 $K(k)$同样是第一类完全椭圆积分，K'则是它的余函数。

最后，下式仍然成立，

---

[15] 这里使用了与 I. Bah 和 P. Bhartia 的 *Microwave Solid-State Circuit Design*（微波固体电路设计）(Wiley, New York, 1988）相同的符号和公式。而他们引用了 G. Ghione and C. Naldi, “Analytical Formulas for Coplanar Lines in Hybrid and Monolithic MICs”(用于混合及单片 MIC 的共面波导的解析公式)，*Electronic Letters*, v. 20, 1984, pp. 179-81，作为最原始的来源。

$$K'(k) = k(k') \tag{22}$$

其中，

$$k' = \sqrt{1-k^2} \tag{23}$$

虽然这些方程很复杂，但它们比较容易编入电子制表软件或类似的工具中。

由于这些边缘耦合结构把电场集中到了横向方向，随着间距和高度之比变小，衬底变得不那么重要了。因而，如果衬底是带损耗的，传输线的衰减实际上可以得到改善。同时，在衬底中电通量的减少也会使有效的非均质性影响降低，从而减轻了色散。

7.3.2　线线之间的不连续性

有时，也会用到不同特征阻抗的传输线。即使只用单一阻抗，直的传输线也不是在任何情况下都是方便(或者说可能)的。这就需要了解传输线段之间的不连续性。例如，把两段特征阻抗不同的传输线(宽度不同)连接起来，相应的电路模型不是简单地等同于两根具有不同阻抗的传输线相串联——这与你所期望的不尽相同。这种复杂性源自伴随着几何失真而产生的场线分布图形的失真。在传输线连接处的附近必须有额外的场分布成分，来满足该处的边界条件。这种附加的场分布成分必然导致了纯的 TEM 传输不再存在，即高阶模一定被激发了。回想一下在推导特征阻抗时假设了 TEM 的传输，那么在这种情况下需要更为复杂的电路模型也就不足为奇了。因为同样的原因，如果传输线存在着弯折，复杂度就会增加，即便是对于一根恒定阻抗的传输线。

这样，纯 TEM 传输的假设不允许在几何不连续处对场分布进行恰当的缝合。一般地说，如果这些不连续性引起了不可传输的模，那么它们的作用主要就是给阻抗增加一个电抗元件。如果能量主要存储在电场中，则该电抗是电容性的；如果电能主要存储在磁场中，则该电抗是电感性的。如果不连续的物理延伸相比所关心的所有波长都较短，则用一个“集总”元件就能充分描述。否则，就需要 π 或 T 形模型(或是这些模型的级联)来进行描述。

首先考虑一个开路传输线这种简单情况。由于边缘效应，在传输线末端，电场并不是突变为零。这种现象比较常见。这也是平行板电容器(以及天线)的表现为比其实际尺寸大的原因。因此，对开路线的一个更为准确的模型实际上是一个带有容性负载的线，如图 7.5 所示。

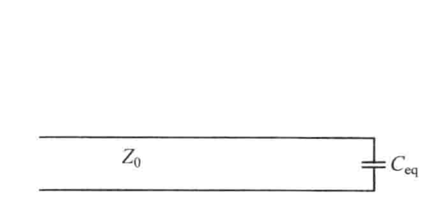


图 7.5　末端开路的微带线的更为实际的电路模型

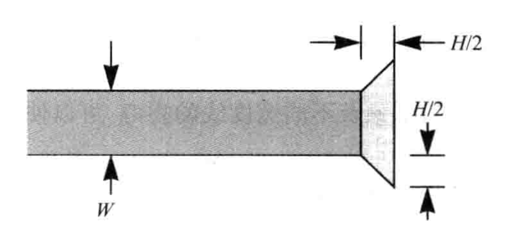


图 7.6　带有末端边缘效应的近似(顶视图)

有效延伸长度的初次估计可以是 $H/2$，其中 H 是前面定义的电介质厚度(如果你真的对大量的公式有兴趣，可参考 7.9 节的推导)。从顶视图(图 7.6)中，可以看到边缘电场也会延伸到传输线的上下，也许会增加另一面积约为 $(H/2)^2$ 的有效区域。这样，进行第二个近似：

$$C_{eq} \approx \frac{\varepsilon}{H}\left(W\frac{H}{2}+\frac{H^2}{4}\right)=\varepsilon\left(\frac{W}{2}+\frac{H}{4}\right) \tag{24}$$

如果需要更为准确的对有效长度延伸 Δl 的估计，可以用 Hammerstad 和 Bekkadal 的经验方程[16]：

$$\frac{\Delta l}{H} \approx 0.412\left(\frac{\varepsilon_{re}+0.3}{\varepsilon_{re}-0.258}\right)\left(\frac{W/H+0.262}{W/H+0.813}\right) \tag{25}$$

这里，ε_{re}是有效相对介电常数，由以下公式给出：

$$\varepsilon_{re}=\frac{\varepsilon_r+1}{2}+\frac{\varepsilon_r-1}{2}\left(1+10\frac{H}{W}\right)^{-1/2} \tag{26}$$

现在考虑一根具有标称固定特征阻抗传输线上的一个拐弯。首先，注意到这种结构在整个拐弯中没有保持固定的线宽(大致可定义为两个相交的边界之间的区域，如图 7.7)。相邻的另外一块在外角上的金属可一阶近似为一个并联电容(更为精确的模型是一种 π 形低通网络，两边是并联电容，中间由一些串联电感连接[17])。在要求不高的应用中，该电容性负载可以忽略。然而，电抗性负载(因而是低通滤波)通常更为棘手，必须加以解决。

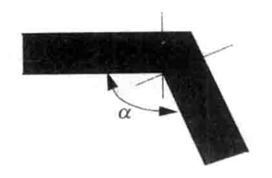


图 7.7　急转角的线拐弯

我们至少需要知道由弯折角区域引入的电容有多大，从而估计其影响。作为一个粗略的近似，这个并联电容可以约等于图 7.8 中的浅色三角形部分。这一结论的逻辑推理是，可以简单地通过切掉这个惹祸的三角金属部分来去除这个影响，如图 7.9 所示。该图所示是一种更一般的情形，其中，斜边长度是 a，因为通过仔细分析发现，当去除的金属面积比图 7.8 浅色区域面积稍大时可以达到最优补偿。给出一个具体的例子，大约 $1.8W$ 可为直角弯折提供最优补偿，而不是$\sqrt{2}W$[18]。所幸的是，最优情况附近相对平坦，稍微偏离最优尺寸也不会带来很大的恶化。

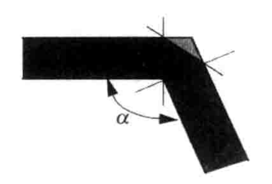


图 7.8　由急转角线拐弯带来的额外电容

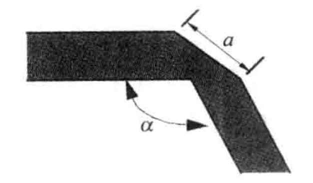


图 7.9　带有斜面的(斜接)线拐弯

为了顾及不同宽度线的拐弯，可以使用图 7.10 中的方法[19]。这里，角 θ 为：

$$\theta=\arctan(W_1/W_2) \tag{27}$$

内角顶点到斜边的距离 b 为

$$b=0.4\sqrt{W_1^2+W_2^2} \tag{28}$$

---

⑯ Hammerstad and Bekkadal，*A Microstrip Handbook*(微带线手册)。

⑰ 串联电感对电流流经弯折时产生的磁能存储效应进行建模。另一个替代办法是，带有两个串联电感和一个并联电容的 T 形网络也是完全可以接受的模型。

⑱ D. M. Pozar，*Microwave Engineering*(微波工程)，2nd ed.，Wiley，New York，1998.

⑲ 此方法取自于 Puff 2.1 user's manual，chap. 7(用户手册第 7 章)。

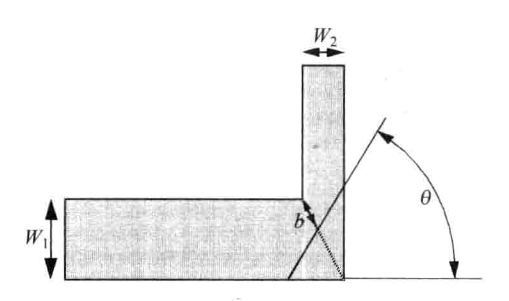


图 7.10 对两个不同宽度的导线形成的直角拐弯进行斜切

因数 0.4 是 Puff 中给出的预设值，但是可以根据需要加以调整。

另一种拐角过渡是光滑的环形弯折，其曲率半径至少应该是线宽的几倍(根据经验，最小一般是 3 倍)。这种弯折可以使寄生电抗减小到可忽略的程度，被称为席卷或环形拐角，如图 7.11 所示。但是斜切拐角更受欢迎，因为这种弯折足够满足大部分应用，大部分版图工具都可以支持。这是因为工程师一般都比较懒(或者说追求效率)。

另一种常见情况是由一个阻抗到另一阻抗(宽度)的阶梯过渡。作为一阶近似，同样的阶梯不连续可用并联电容来建模，如图 7.12 所示。就像对所有这些类似的情况，描述阶跃不连续寄生效应的方程往往是十分复杂的。然而，直观上(粗略的)可以近似认为：较宽的导线产生电容的不连续。该电容可由一阶近似计算得出，类似于面积为 $l(W_2-W_1)$ 的平行板电容，其中 l 是宽导线的长度。若考虑了边缘电容，还可以对这个估计进行改进，但这仅对小 l 的电容才需要。以上这两种情况的一阶模型包括两段宽度为 W_1 的传输线，它们之间有一个接地的并联电容。

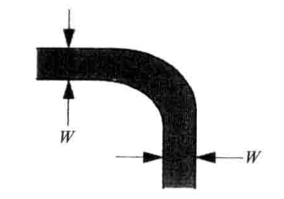


图 7.11 环形(或横扫的)拐角

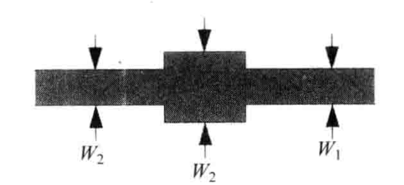


图 7.12 阶梯状的不连续

当连接两根不同宽度的导线时，最好使用楔形过渡而不是阶梯过渡。渐变楔形可以减小存储在高次模中的电抗性能量密度。为了达到“充分渐变”，有一个粗略的原则，就是使楔形区的长度至少是最低频率分量的四分之一波长那么长，而这个最低频率是要有好的匹配的。虽然图 7.13 给出了一种简单的线性楔形(为画图简便起见)，如第 4 章所阐述，其他楔形形状会更好些。

前面的所有这些例子很容易使读者认为不连续的电学效应本质上都是电容性的。为了使读者摒弃这种想法，图 7.14 给出了一种电感性寄生的情况。既然现在有了更一般性的答案(指不仅是电容性的，译者注)，可以假装早已预见到所有这些情况。狭窄部分会产生电流拥挤，从而增加该处的磁通密度。这种增加可以解释为一种电感性效应。

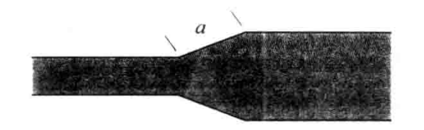


图 7.13 两条线之间的楔形过渡

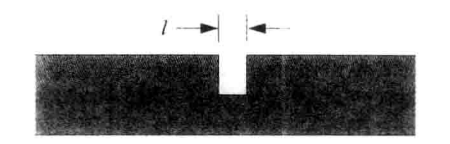


图 7.14 带有槽口的传输线

换一种说法，考虑如下的情况：如果宽度较小的部分是无限长的，那么波的传播是 TEM 型，且其阻抗比线宽大的部分大。这样，对于任意一根以较低阻抗终结的适当短的导线，其阻抗行为会类似于电感。即使我们知道在导线宽度阶跃变化位置附近，TEM 传导是一个很差的假设，仍将在该假设下计算电感值。

从一段传输线的阻抗计算公式（见 7.4 节的公式(31)），在我们已声明的假设条件下，可以推出短传输线段电感的表达式：

$$L \approx \frac{lZ_{0,n}\sqrt{\varepsilon_{r,eff,n}}\left[1-\left(\frac{Z_{0,w}}{Z_{0,n}}\right)^2\right]}{c} \tag{29}$$

其中，n 和 w 分别代表窄和宽。这个方程至少是具备了正确的大致趋势：即当槽口深度趋于零时，电感也趋于零。由于推导此方程时用了些未经严格证明的假设，除非是用于粗略的估计，一般不要依赖于这个方程。

更细致的模型，不仅考虑了之前讨论过的，由同样的长度延伸机制导致的并联电容（末端到地的边缘效应），而且还考虑了与电感并联的电容，该电感产生于传输线末端到末端的边缘效应。这些改进模型留给读者作为练习。

另一种常见的几何形状是传输线 T 形连接。在这种情况下，两根导线在连接结点处两个内角附近的电容耦合有效地构成了一个旁路。如图 7.15 箭头所示，这个电容耦合（大部分来说）实际上在转角处形成了短路，使得与水平导线相接的一小段垂直导线变得没有关系。也就是说，这一小段好像不存在一样；垂直导线表现为比实际导线短的性质。很多情况下，这种缩短并不是个问题。但也有一些情况需要特定的长度。对于这些情况，必须增加导线长度，从而使其具有正确的电长度。一阶的估计可以认为导线应当增长约 H，即电介质厚度。方程(61)（7.6 节）提供了更精确的需要增加的长度的计算。

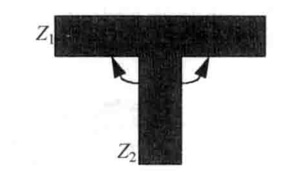


图 7.15 T 形连接的短路效应示意图

就如在与 T 接头那样，缩短效应也会因为同样的原因出现在宽窄导线的接头处。缩短量 Δl 由以下公式近似给出[20]：

$$\frac{\Delta l}{H}=\frac{120\pi}{Z_1\sqrt{\varepsilon_{re,ser}}}\left[0.5-0.16\frac{Z_1}{Z_2}\left(1-2\ln\left[\frac{Z_1}{Z_2}\right]\right)\right] \tag{30}$$

7.3.3 连接器和传输线之间的过渡

从一个同轴电缆连接件到平面传输线的过渡如果处理不当会引入严重的阻抗不匹配现

[20] *Puff* 2.1 manual（手册），p. 37. 该手册标注的原始出处是 Hammerstad and Bekkadal, *A Microstrip Handbook*（微带线手册）。

象。就像传输线与传输线间连接的情况，一个重要的要求就是要避免急的拐角，以减小对阻抗的影响。然而，严格满足这个要求通常意味着要使连接器的轴芯导体与电路板在一个平面上(图7.16)。毋庸置疑，这种电气性能最优的连接方式，在很多情况下都是不方便的，因为它把连接器的位置限制在了印刷电路板的外围。

在印刷电路板的表面上安装连接器会方便很多，但这使信号不得不通过一个直角弯折。通过印刷电路板的斜接或环形过渡都不易实现，这样看来信号质量的恶化在所难免了。尽管有些恶化不能避免，但其幅度可以通过以下一些版图设计规则大大减小。这些补偿方法可以从我们已经研究过的纯微带线的经验中预计到。就像早先的平面情形，我们期望直角弯折引入一个并联电容。补偿涉及在弯折附近对版图的修改。以SMA连接器为例。在中心管脚周围的大部分接地平面都需要去除，以减小并联电容，从而在过渡中保持阻抗近似恒定，如图7.17所示。其他不同的要安装到印刷电路板上的连接器都需要遵守同样的版图规则。



图7.16　架在PCB板上的连接件(侧面图)

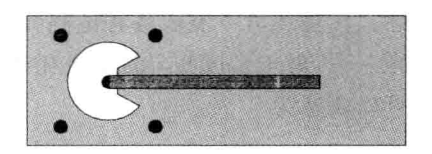


图7.17　改进型的SMA到微带线的过渡(从PCB板带有走线的那面看下去)

注意SMA和BNC类的印刷电路板常需要钻五个孔(一个为了中心导体，其余四个是用于固定接地的法兰圈)。为了快速的样品制作，可以采用带有底盘安装的BNC，这样只需要打一个孔。但也有一个挑战，因为这种BNC螺纹衬套要比印刷电路板的厚度长。这样，如果钻一个适合衬套的孔，中心引脚会突出电路板表面以外很长距离。然而，如果使用一个非正规安装方式：钻一个能容纳BNC聚四氟乙烯电介质的孔，而不是螺纹，只使这部分伸出电路板。然后以这样的方式安装BNC，即把螺纹焊接到电路板上(所需的热量可能超过许多小的烙铁可以提供的热量)，在焊接周沿时要小心，整个一圈都要焊好，并使焊缝均匀[㉑]。将微带线与中心引脚焊接在一起就完成了，中心引脚应当与电路板取平(用1/16英寸，或1.6mm厚的材料，以达到近乎完美的配合)。如果需要的话，要修整连接器的电介质，使其也与电路板齐平。显然，这种不专业的安装方法完全不适合商业产品，但对于业余爱好者和学生在几个GHz频率范围内的应用还是非常有效的。不管怎样，这种方法是与FR4具备的品质相配的，并且是迄今在上述频率范围内最低成本的电路制作方法。

尽管听起来十分不专业，这种方法在过去二十年中，在斯坦福大学的微波电路实验室，都有很好的测试结果，如图7.18所示。如果使用了上述非正规的同轴到微带过渡，你会发现，在直角连接处总有一点电感性的不连续性，说明这种方法过度补偿了典型的电容性弯折。在过渡连接处附近增加一点电容(用适量的铜箔胶带)，会大大改善连接的性能。如果连接过渡的质量很重要，那么强烈推荐在网络分析仪和时域反射仪(time domain

㉑　确定连接器表面不是镀以不锈钢的，不锈钢的不能焊接！在这些情况下，可以用砂纸或锉刀夹去除掉妨碍焊接的表面层。

reflectometer，TDR）的引导下进行反复的调试。

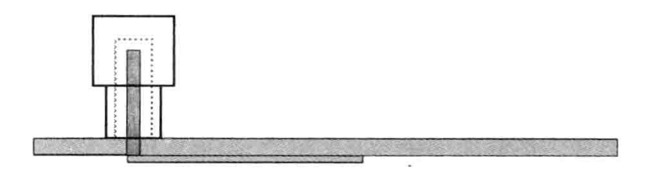


图 7.18　非常规的对于带套筒的 BNC 连接件的安装方法（侧面图）

7.4　用传输线段构成的无源器件

在本节中，我们明显地指明传输线在使用中是一个阻抗变换器。这种变换源自于和有限传播速率相关的电压和电流延迟。V 和 I 之间相移的长度相关性会相应地引起阻抗的长度相关性。传输线可以制成各种电路元件就是对这种现象的利用。

回顾一下，对于长度为 l 的无损耗线，接任意负载阻抗 Z_L，在传输线 $z=0$ 一端看进去的阻抗可以表达为

$$\frac{Z(l)}{Z_0}=\frac{Z_{Ln}+\mathrm{j}\tan\beta l}{1+\mathrm{j}Z_{Ln}\tan\beta l} \tag{31}$$

其中，β 是传播常数 γ 的虚部，

$$\beta=\mathrm{Im}[\gamma]=\omega\sqrt{LC}=\omega/v \tag{32}$$

其中，v 是传播速率。对于微带线，后者的计算较为复杂，因为微带线周围的电介质不是均匀的。因此，有效介电常数介于电路板材料和空气介电常数之间。有效介电常数的其他近似计算公式如下[22]：

$$\varepsilon_{r,\mathrm{eff}}\approx\begin{cases}1+0.63(\varepsilon_r-1)\cdot(W/H)^{0.1255}, & W/H\geqslant 0.6 \quad (33)\\ 1+0.6(\varepsilon_r-1)\cdot(W/H)^{0.0297}, & W/H<0.6 \quad (34)\end{cases}$$

对于典型的 FR4 上的 50Ω 线，$W/H\approx 2$，因此有效介电常数大约在 3.5（若实际体介电常数 ε_r 是 4.5）。这样，波长比在空气中的一半稍长。

从物理和直觉角度，我们都可以期望短的传输线段可以用于实现电感和电容。来看两个极端情况。首先，假设传输线以开路终结，那么这样的一小段就是一个电容，因为由电介质隔开的两个导体就构成电容。相反地，如果一小段传输线以短路终结，就会表现出电感性，因为现在我们有了电流回路。

为了推导出电容的显式公式，令方程(32)中的负载阻抗趋于无穷大。一小段开路线的输入阻抗因此近似为

$$Z\approx\frac{Z_0}{\mathrm{j}\omega(l/v)} \tag{35}$$

因此，电容为

[22] Carson, *High Frequency Amplifiers*（高频放大器）. 据称，这些方程不及 Hammerstad Bekkadal 早先提出的方程（式(26)）。

$$C = \frac{l}{\nu Z_0} = \frac{l\sqrt{\varepsilon_{r,\mathrm{eff}}}}{cZ_0} \tag{36}$$

一个有助的记忆方式是时间常数 Z_0C 恰好是单程的波传播时间。

对于空间不是那么紧张的应用中，不应当忽视这种制作电容的方法。即使不是所有的电容都用这种方法制成，为调整方便起见，至少有一部分电容应采用上述方法，这样在电路板制作好以后还可以调整。在任何情况下，知道给定面积的电容器的电容值是多少是有价值的，哪怕仅仅是为了估算版图的寄生参数。例如，对于 FR4 材料，衬底厚度为 1/16 英寸时，单位面积电容大约是 2.5 pF/cm^2。另外，FR4 的损耗比分立电容中使用的 X7R 或 Y5V 陶瓷电介质材料的损耗稍小，但在同一范围内。当然，使用 PTFE 或 RO4003 等较高质量的电路板材料，能够使损耗更小(单位面积的电容值也会稍小)。

要特别注意，电路板的尺寸要比我们所关注的最高频率波的波长要小。否则，分布效应会改变阻抗特性。

类似地，对于以短路终结的传输线段，其电感值为

$$L = \frac{lZ_0}{\nu} = \frac{lZ_0\sqrt{\varepsilon_{r,\mathrm{eff}}}}{c} \tag{37}$$

我们再次发现，时间常数(这里是 L/Z_0)是单程信号波传递时间。

其他各种电感公式在附录 A(7.8 节)中给出。其中许多已经在先前的第 6 章出现过，但为了参考方便，又在附录中整理出来。

器件一旦设计好，我们就要确认计算出的传输线长度与电学波长(可由已给出的有效介电常数计算公式)相比是非常小的。否则，这种近似会有很大的误差。

在实践中，这些电感和电容可能不总是接到开路或短路。为了使前面这个要求得以保证，可以控制传输线的特征阻抗，这样，它们的末端就可以是开路或短路(见式(29))。所以，制作电容时，得使其 Z_0 尽可能低(或根据实际情况可能达到的低值)。反之，制作电感时，Z_0 要尽量大，以使得任何可能接到该传输线的阻抗负载，相对来说它都可以近似为短路。

除了可以制作其值在一个连续区间的元件的灵活性外，传输线可以使我们得到性能比分立器件更好的器件。例如，制作 1 nH 的电感，用微带线来实现要比用分立元件容易得多。

用传输线制作元件时，有些实现的问题需要考虑。首先，用作电感的传输线的特征阻抗不可能是任意高的，因为制造可靠的线宽总有一个最小值。假定一种典型的制造公差是 2mil，并且认为这个偏差被允许最多到总线宽的 20%，这样可制作的最小实际线宽是约 10mil[23]。因此，在 1/16 英寸厚的 FR4 上，实际导线特征阻抗很少超过 200Ω。但这个值对于许多用途来说已经足够高了。然而，这个有效的串联电阻看起来有点偏大，所以这种电感不能用于需要高 Q 值的场合。

制作电容时，有两个实际约束条件。如果线宽太大，就与波长可比，从而使近似失效。另一个约束就是电路板的空间。

在下节中，我们放宽尺寸上的限制，也会考虑损耗小的传输线。这样，我们推导出了用

---

[23] 这些值是 FR4 的典型值。在更硬的衬底上，可能实现更窄的线宽。依照衬底的类型，也许 1mil(25 μm)的线宽代表了目前工艺水平，即使这么窄的线不常见。

传输线构制的元件有效串联电阻的显式解析表达式，从而可以计算 Q。

7.5 谐振器

前面这一节假设了这么一个约束条件，即所有的传输线段长度与波长相比都要小得多。然而，式(31)不受这个条件的限制；对长度的限制只是保证为了只有纯电抗生产。本节，我们讨论违背上述约束的情况，从而产生“不纯的”电抗。具体地，我们来看一段标称电学长度是 $\lambda/4$ 的短路线的输入阻抗。在使传输线段的电学长度恰好是 $\lambda/4$ 的频率上，理想的输入阻抗是无穷大(假设没有损耗)。随频率上升，该传输线段的电学长度会大于 $\lambda/4$。因而，输入阻抗呈电容性(可用式(31)验证)。在有效电学长度小于 $\lambda/4$ 的频率上。输入阻抗表现为主要是电感性。从上述描述可以清楚地看出，这样一个传输线段的行为非常像并联 RLC 网络，甚至在 DC 时能提供通路这一点上也相像。但两者有一个重要的不同点：传输线段有多个谐振点，而集总电路只有一个。

对一个开路的 $\lambda/4$ 传输线的对仗分析表明：其行为与串联 RLC 网络非常相似。不论开路还是短路，$\lambda/4$ 传输线的谐振特性已经广泛地应用于振荡器和滤波器中(分别在第 15 章和第 23 章中详细讨论)。

对于这类振荡器，Q 值是导体与电介质损耗以及辐射的函数。因其重要性，现在进行这个关键参数的近似推导。考虑到我们已经发现了它们和 $\lambda/4$ 长传输线之间的相似性，我们会运用在集总 RLC 网络学到的知识。

对于一个集总、串联谐振网络，谐振点处的 Q 值可以表达为电感性(或电容性)电抗与阻抗的实部之比。对于一个带损耗传输线，归一化阻抗为

$$\frac{Z(l)}{Z_0}=\frac{Z_{\mathrm{Ln}}+\tanh\gamma l}{1+Z_{\mathrm{Ln}}\tanh\gamma l} \tag{38}$$

其中，传输线的零点位置仍看作在负载一侧，坐标 l 是阻抗测量端口的位置。

在我们这种情况下，负载阻抗是无穷大，因而，标准输入阻抗为

$$\frac{Z(l)}{Z_0}=\frac{1}{\tanh\gamma l}=\frac{1}{\tanh(\alpha+\mathrm{j}\beta)l} \tag{39}$$

因为可用的任何谐振器的损耗都不能太大，可假定有 $\alpha l\ll 1$，从而简化推导。满足这一条件后，谐振时的输入电阻近似为

$$R_{\mathrm{eff}}\approx Z_0\alpha l=Z_0(\alpha\lambda/4) \tag{40}$$

我们已经得到了一段短传输线的等效电感表达式，但这里不能用，因为这里的传输线没那么短。相反，通过将传输线阻抗斜率与相应的集总 LC 网络的斜率等同起来，就会发现这里恰好有两倍的因数差异：

$$L=\frac{Z_0\pi}{4\omega_0} \tag{41}$$

从能量角度考虑，有效谐振电感要小两倍因数是可以预计到的。电流以正弦方式变化，而不是常数，因此，磁场中存储的能量是假设电流为常数时的一半。

由这个有效电感表达式，可以计算出开路 $\lambda/4$ 传输线的品质因数：

$$Q=\frac{\omega_0 L}{R_{\mathrm{eff}}}=\frac{\pi}{\alpha\lambda}=\frac{\omega_0}{2\alpha v_{\mathrm{p}}}=\frac{\omega_0\sqrt{\varepsilon_{\mathrm{r,eff}}}}{2\alpha c} \tag{42}$$

这些关于 Q 的表达式是非常通用的，允许我们从测量或计算到的传输线衰减和速率因子中推出可以达到的 Q 值(反之亦然)[24]。

举个具体的例子，用之前提及的 FR4 的数据。50Ω 传输线的衰减大约是 0.07dB/(in · GHz)，相应于每 GHz 的 α 值约为 0.35 Np/m。[25] 在 1GHz 下，0.35 Np/m 的衰减意味着一个无负载的谐振器的 Q 值约为 50(典型的有负载 Q 值是这个值的一半)，给定典型相对介电常数为 3.5[26]。由于 FR4 材料的传输线衰减在一个很宽的频带里都与频率成比例，当使用这种板材时，可实现的谐振器的 Q 值随频率变化很小。若需改进，但又不能更换衬底(比如，受限于成本)，那么将高品质无源元件与微带线一起使用是明智之举。例如，如果需要改进一个由开路微带线制成的谐振器，我们可以使其变短(使它更具电感性)，然后，可在该微带线末端和地之间接一个电容，进行补偿。缩短传输线会减小其上的电压峰值，从而减小介电损耗。通过有选择地使用高质量的分立元件来对价格低廉的微带线进行补充，可以得到性能极佳的电路。

将 FR4 的无负载谐振器的约为 50(这当然是一个相当有用的大的值)的 Q 值与其他材料和结构可以实现的 Q 值进行比较是十分有指导意义的。若使用 RO4003，可以将 Q 提高四倍，因为该材料有非常小的介电损耗。使用其内在电介质低损耗的同轴电缆可以使 Q 值再提高三倍，这样就可以实现具有 Q 值在 1000 数量级的谐振器。谐振腔还可以使 Q 值再增大 10 倍，这种谐振腔可以是由导体做成的闭合的盒子或者是由特殊电介质材料制成的固体块。Q 值的这种不断上升可以定性理解如下。暴露在外的平面结构(如微带线)的 Q 值最低，因为除了导体和电介质损耗外，还有辐射损耗。同轴结构比微带线好，因为这种结构可以屏蔽辐射。谐振腔更好，一方面是因为其较低的表面积与容积比，另一方面是因为没有了轴芯这个耗散性导体。前者的考虑是重要的，这是因为粗略来讲，能量存储在体积内，而在表面耗散掉。一个立方体(而球体更好)比圆柱体有更好的表面积容积比。介电谐振器没有任何耗散性导体，若用一种适当的低介电损耗材料，就可能得到极高的 Q 值(几万)。

7.6　合成器　分配器和耦合器

经常会有需要把从多个信号源发来的信号输入整合成一个输出的情况。例如功率放大器，需要把几个低功率级整合起来，得到一个高功率的输出。对于其他线性无源元件，这些合成器可以反过来用作功分器。作为功分器时，可以得到微分或正交的输出，这取决于设计细节。

由于这些元件具有这些用处，工程师们已经花了大量的努力来设计各种实现方式。本书篇幅的限制迫使我们只关注更为常见的类型。我们也给出了集总和分布式的版本，这部分是为了帮助我们用一种实现方式作为另一种方式的逻辑扩展，促进相互理解，部分是为了扩充我们的选择余地。

---

[24] 一个经常被问及的问题是："为什么 Q 总是用电感来表达？"回顾一下，在谐振时电容与电感的电抗是相等的。这样，我们如果用电容来计算 Q 值，我们得到同样的结果(不要只听我说，你自己可以算一下)。之所以把注意力集中到电感器仅是因为历史上的原因：分立的电感器的 Q 值通常比电容器的低，因而工程师总是更关注电感器的 Q 值。

[25] 记得一个奈培(neper)是 e 倍。幅度变化一个奈培因此对应于功率约变化 8.69dB。

[26] 回想一下有效(介电)常数计及了微带线非均质介质这一情况。

7.6.1 电阻性合成器

最简单的分配器/合成器是由一个三电阻网络构成的，如图 7.19 所示。根据图上标出的数值，读者可以自行验证：若所有端口都以 Z_0 终结，网络就会匹配。同时，这种情况有 6dB 的衰减。但作为这个损耗的补偿，电阻式功分器在很宽的带宽上都能很好地工作。

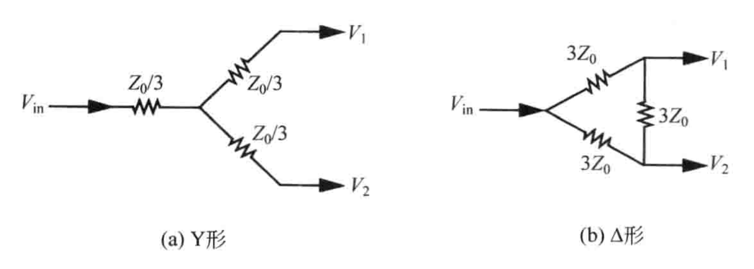


图 7.19　用分立元件实现的电阻型 6dB 分配器与合成器

7.6.2 分布式合成器

如果合成器具有衰减是不能被接受的（不能接受合成器有衰减），还有其他替代方案。最简单的无损耗微带线功分器/合成器由三条微带线组成的结构，其中两条的特征阻抗相等，并且都是第三条的两倍，如图 7.20 所示。

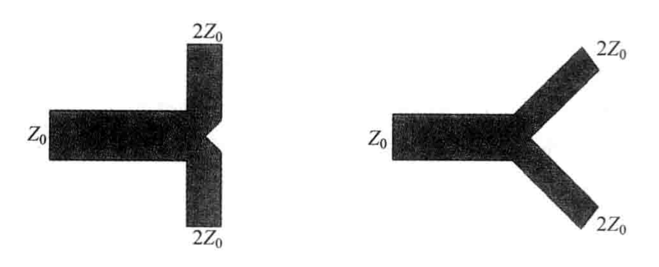


图 7.20　T 形结与 Y 形结分配器与合成器

由于每个臂上需要以它的特征阻抗终结，这种专门的合成器在三个端口上需要不同的终结。这个性质是一个潜在的缺点。T 形合成器在三个臂的连接处也存在严重的阻抗不连续性。如图 7.20 所示，为了对不连续引起的并联电容负载进行一阶补偿，通常需要在垂直导线上对着 Z_0 臂处切出一个三角槽。Y 形合成器的这个问题不太严重，因其没有那么突然的过渡。我们还可以用弧线过渡（而不是图上给出的急过渡）来获得更好的性能。

Ernest J. Wilkinson（威尔金斯）给出了一种非常巧妙的 Y 形结修正，可以使各个端口匹配，同时输出端口间互相是隔离的[27]。对于最常见的应用，功率可以被分配到任意数量的端口上去。我们首先考虑最简单的情形，将功率等分到两个端口上。

如图 7.21 所示，有三个明显的端口。我们还可以看到实际上存在着第四个端口，它位于电阻的中间点。因为这个端口一般不用作输出端，所以很少明确指出。然而，有时考虑它

---

[27] Wilkinson，当时还在 Sylvania 火箭系统公司（译者注），对这个结构在“N-Way Hybrid Power Combiner”（N-路混合功率合成器），*IRE Trans. Microwave Theory and Tech.*，v. 13，January 1960，pp. 116-18，进行了描述。

的存在是非常有用的。当我们考虑 Wilkinson 分配器的通用化时，就会看到这点。

为了确定对应于端口隔离和等功率分配条件的 Z 和 R 的值，可以利用这个结构的对称性，并且通过共模和差模输入来探讨这个合成器的工作原理。由通常的电路理论可知，任何一对电压(V_1,V_2)或电流都可以用它们的平均值(共模)和它们的差值的叠加来表示，如图 7.22 所示。在微波电路中，习惯于采用偶次模和奇次模激励(而不是共模和差模)。它们具有相同的意义，只是语言表述习惯不同。不论何种提法，Wilkinson 合成器对任意输入电压对的响应，都可先分别计算共模和差模响应，再把这两种响应加起来得到。

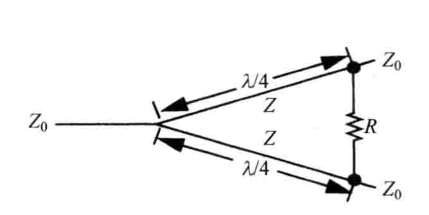


图 7.21　Wilkinson 分配器/合成器(形式化的图)

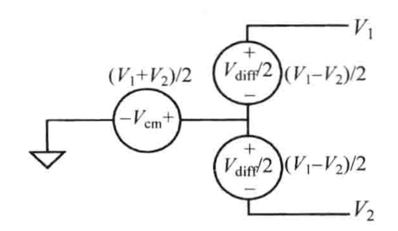


图 7.22　将任意的一对电压分解成共模与差模分量

首先考虑共模响应，假设由相同电压作为合成器的输入。这时，由对称性可知，合成器上下两半的行为相同。为了强调这种对称性，可将电路简略画成图 7.23。

我们所要做的只是把单个输出线换成带有两个具有两倍阻抗的传输线的并联组合(注意顶点处的短接线)。由对称性可知，不会有电流跨越一根假想的平分整个结构的水平线。因而，可以将上下两半电路分开[28]。因为没有电流流过电阻，所以可以将其整个去掉。类似地，短接两根双阻抗值线的小的垂直线中也没有电流。这样将它们整个分离开，就得到了一个简单的电路，在这个电路中，两半电路都能够完整地传递电路的偶次模响应。如果我们去掉多余部分，如下半部分电路，那么剩下的电路如图 7.24 所示。显然，如果我们使 Z 等于输入和输出端阻抗的几何平均，就可能对偶次模进行阻抗匹配：

$$Z = \sqrt{(Z_0)(2Z_0)} = \sqrt{2}Z_0 \tag{43}$$

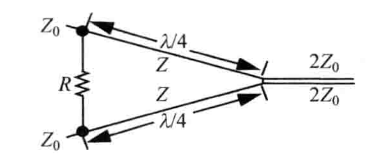


图 7.23　Wilkinson 合成器(依然是形式化的)

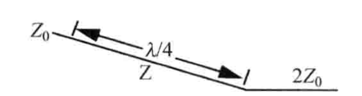


图 7.24　经过了对偶次模激励简化后的 Wilkinson 合成器

在用偶次模得到了所需的臂支路的特征阻抗后，然后用奇次模来分析，希望借此来得到关于 R 的信息。这种情况下，一个纯差分(奇次模)激励会产生不对称的响应，再一次可以使我们进行简化。因为此时上下两半电路中的电压互为镜像，从整个电路结构中间穿过的

[28] 另一种等效的方法是将上下两半电路并联起来，因为在对应的镜像点上在共模激励时的电压相等。用这种方法也会得到相同的答案。

水平线是电压为零的一条等值线，即虚地。简化后，合成器的半边电路如图 7.25 所示。这里我们可以看到，高阻抗臂的顶点是接地的。当由 $\lambda/4$ 支路变换到输入段时，这种短路就变成了开路。那么，奇次模输入时的唯一负载是由电阻提供的。为了匹配奇次模成分，只需满足：

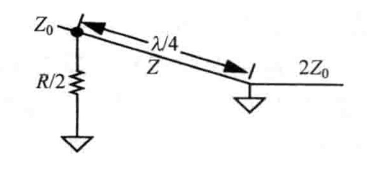


图 7.25　经过对奇次模激励简化后的 Wilkinson 合成器

$$\left(\frac{R}{2}=Z_0\right)\Rightarrow R=2Z_0 \tag{44}$$

就像其他对称电路（如差分放大器）一样，可以看到，偶次模和奇次模分解法可使分析与设计大为简化。在这种情况下，可以对奇偶次分量分别匹配，然后再满足上述条件。最终，这种分析方法表明，Wilkinson 合成器确实可以提供隔离，就像开始时我们提到的那样。我们已经看到，对于偶次模激励，上下两半电路互相去耦，因为两者间的连接都等效于开路。对于奇模电路，上下两半电路还是可以互相去耦的，这次是因为在桥接电阻中点和支路顶点都形成了虚短路。由于任意激励都可以表示为偶模与奇模之和，这种去耦方法具有一般性。上述这些特性使 Wilkinson 分配器/合成器成为了非常受欢迎的微波元件。

我们的分析表明，桥接电阻只是对于奇模元件起短接作用。一般情况下，Wilkinson 分配器/合成器是对称工作的（即处于纯偶次模）。因此，桥接电阻仅仅是来消耗因在两半电路所连接的负载不是非常匹配时产生的入射的奇次模能量。但这个入射的奇次模信号不一定就是那么小，所以很有必要计算最坏情况下的耗散，然后选择具有相应功率承受能力的电阻。一个常见的错误就是省略这个计算，直到闻到焦味时，才发现电阻的额定功率不足。

实际的 Wilkinson 分配器/合成器版图改进了最初不太理想的地方，如图 7.26 所示。现在我们已经知道，在实际中，到输出端的过渡通常是比较平滑的。在实际中，Wilkinson 合成器的整个结构都常采用曲线过渡（如图 7.27）。

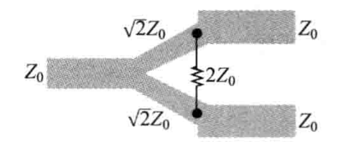


图 7.26　Wilkinson 分配器/合成器（形式化的，很明显不是优化了的）

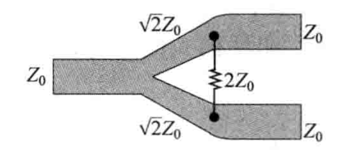


图 7.27　Wilkinson 分配器/合成器（更具有实际版图的代表性）

相同的 Wilkinson 合成器可以级联，从而提供更多的输出。在功率放大器模块中，整合几个低功率放大器的输出时，有时会用到这种二元（或集团 corporate）阵列。Wilkinson 的原始论文还表述了他的方法的另一种一般化，就是分配器可分成任意路（不仅是二进制值）。这种方法也可扩展到不等功率分配。这些可能性能通过图 7.28 这样的直观图示进行很好的理解。

对于双工情况，寻找偶模阻抗匹配可以使我们推导出所需的 $\lambda/4$ 支路阻抗：

$$Z=\sqrt{(NZ_0)(Z_0)}=Z_0\sqrt{N} \tag{45}$$

为奇模元件提供匹配需要在每条线到公共点（之前我们提到过的虚拟点）之间用一个阻值为 $R=Z_0$ 的电阻连接，该公共点上没有其他连接。

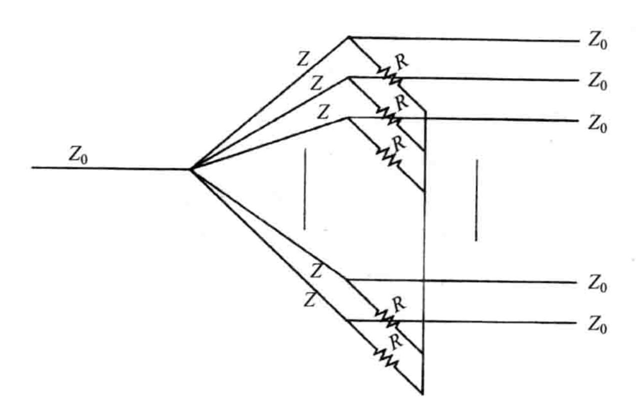


图 7.28　一般化了的 N 路 Wilkinson 分配器/合成器

如果 N 超过 10，所需的导线阻抗可能会升高到实际微带线难以达到的值。另一个实际的考虑是，$N>2$ 时，需要连接桥接电阻，因为简单的串联连接已经不能满足要求了。为了消除或减小这些困难，对于单独或 N 路的应用，有时二元树还是比较可取的。

也可以用一般化的等功率分配器作为设计不等功率分配器的起点。例如，想以 $M:1$ 的比率分配输出功率，可以从 $N=(M+1)$ 的一般化等功率分配器开始来考虑这个例子，如图 7.29 所示。

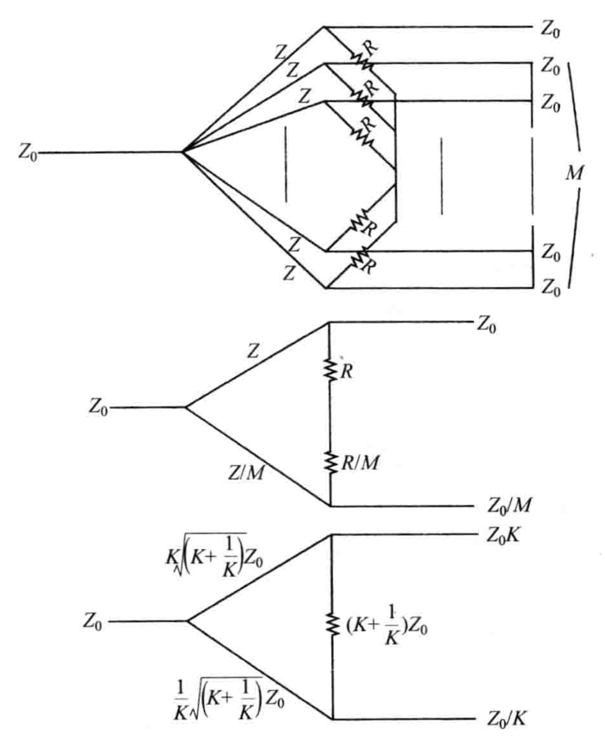


图 7.29　从 N 路 Wilkinson 到不相等的分配器/合成器的演变

在图中演进序列的第一步，我们可以简单地将$(M+1)$中的M个输出接到一起。为使输出匹配，这里，负载阻抗必须是Z_0/M。通过计算余下的高功率支路的并联组合，进行适当的简化后，我们得到了三个设计中的第二个，如图7.29所示。

下面我们对结果进行缩放，使负载阻抗的几何平均等于Z_0。这种归一化减小了实际中窄导线阻抗过高的可能性。这样，输出端接以因数K缩放，支路的阻抗也类似地进行缩放。功率比M就等于K^2。最后，奇模端接的两个电阻整合成为一个电阻。

显然，也可将功率不均等的分到其他输出端。只需从归一化的N路分配器开始，以同样的方式进行。

Wilkinson合成器也可通过集总方式实现（如图7.30）。这种替代方法在调整时更容易控制（只要工作频率不是很高），并且可以更为紧凑，在低频时尤其如此。它与它的微带线共用有限的带宽。如图所示，集总法只是简单地将微带合成器的$\lambda/4$支路替换为一个简单的π形网络传输线。这里，我们调整了第2章的相关设计公式。

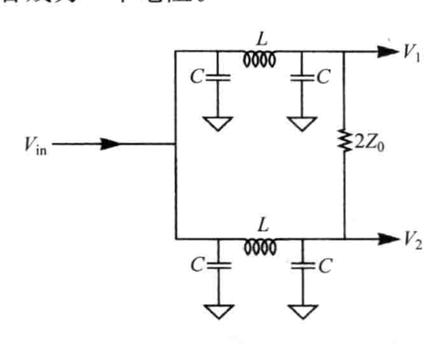


图7.30 集总元件实现的两路Wilkinson分配器/合成器

首先，选择L/C，得到正确的支路阻抗：

$$\sqrt{L/C}=\sqrt{2}Z_0 \tag{46}$$

下面用LC，得到正确的中心频率：

$$1/\sqrt{LC}=\omega_0 \tag{47}$$

从而得到

$$C=1/\sqrt{2}\omega_0 Z_0 \tag{48}$$

$$L=\sqrt{2}Z_0/\omega_0 \tag{49}$$

那么，每个元件的阻抗模值就是相应导线的特征阻抗。注意到两个输入电容也可以合成一个，因此，实际只用到三个电容。图中画出了四个，只是为了便于解释。

最后，电感之间必须有足够的空间，从而使它们间的磁耦合可以忽略。如果两个电感是互相垂直的，或者已经采取了屏蔽措施（如弯成螺旋管），那么可将它们放置近一些。

7.6.3 混合型变换器和巴伦变换器[29]

混合型变换器是非常多种类的元件。首个混合型变换器是为了解决电话通信中的一个难题而发明的。这个课题就是如何可以实现在仅一对电线上实现双工通信。实现双工通信的最直接的方法需要三条或四条电线（一条发送，一条接收，以及它们相应的回路地线）。多亏有了混合型变换器，用一对电线就足以实现完全的双工通信。这个功能的实现依赖于信号可以被分解为共模（偶次模）和差模（奇次模）成分的可能。比如说，通过用偶次模承载信号的发送，用奇次模承载信号的接收，就能在一对电线上实现双工运作，如图7.31所示。

---

[29] *Balun* 源自"balanced-to-unbalanced"，因而与 *gallon*（加仑）谐音。

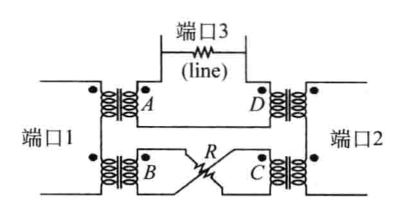


图 7.31　经典的电话通信混合型变换器[30]

加在端口 1 上的电压在变压器的次级线圈 A 和 B 上产生相同的次级电压。端口 A 两端的电压产生出流入端口 D 的电流。相应地，端口 B 两端的电压驱动端口 C，但产生的电流极性相反。端口 D 和 C 上的电压值是相同的，但极性相反。如果你能跟上这个过程，并且能搞清楚极性，那么就可以看到端口 1 上的电压最终出现在端口 3 桥接的电阻上了。即端口 1 耦合到了端口 3。同时，端口 D 和 C 上的电压幅度相等极性相反意味着端口 2 上没有电压。端口 1 和 2 是相互隔离的。因此，这种变压器的连接方式神奇地使得麦克风和耳机能分别接到端口 1 和 2 上，来共用端口 3，但不会产生不希望的内部反馈回路[31]。

读者从以上描述中也许能够想到，基于对称性进行区分的一般想法可以通过好多种方式实现，因为有很多种电路可以对差模和共模信号给出不同的响应。在经典的电话混合型变换器上，就是由缠绕在软铁芯上的多抽头宽带变压器来实现这种区分的。在微波频率下，经典变压器的不理想性促使了人们采用其他的替代方案，但它们的基本原理都是相同的。

在继续讨论混合型变换器前，最好是先给出它的一个定义。hybrid 这个词的词义随时间而改变，但现在它一般指的是任何具有以下性质的四端口元件：所有端口阻抗都是匹配的；至少有一个隔离的输出端口（即在某种输入条件下，该端口为零输出）；并且可以等功率分配。因此，混合型变换器通常是由输出端信号之间的专门的相位关系来分类的。

在许多种混合型变换器的符号中，图 7.32 显示了两种。左图给出了对隔离端口的内部终断。为方便起见，很多混合型变换器都是这样构造的。在右边的简图中，对角线标明相互耦合的一对端口，而不涉及隔离端口是如何终断的这个问题。

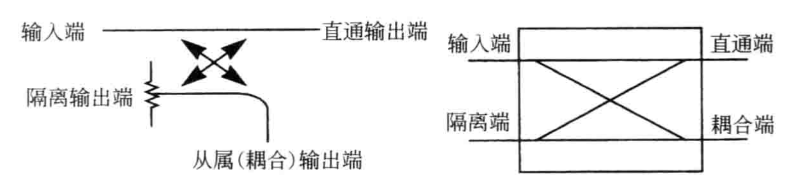


图 7.32　一个混合型耦合器的一些符号

本节中我们会看到，存在着许多种混合型变换器，它们之间的区别在于具体的实现细节以及端口信号之间的相位关系。一个 180°混合型变换器，之所以这样命名是因其能提供两个反相的输出，可实现单端到差分的变换（反之亦然）；可用作分配器或合成器（但并不是所有的分配器或合成器都是混合型变换器）。对于平面微波实现，窄带 180°混合型变换器由

---

[30] 这种巴伦变换器是用于拨号电话机的几种变换器中的一种。

[31] 在实际的电话中，故意让这种端口隔离不太完美（通过选一个与导线阻抗不同的抵消阻抗 R，见图 7.32），这样说话人就能听到自己的声音。为这种侧音实现的声音反馈可以帮助说话人使用合适的音量。

一个电周长是 $3\lambda/2$ 的闭路（如环形回路）微带线构成。四个抽头（图 7.33 中 A,B,C 和 D）的彼此间隔为 $\lambda/4$。加在 A 上的信号分成顺时针和逆时针两路。信号从 A 到 B，顺时针走向的信号变化了 $\lambda/4$，而逆时针走向的信号则变了 $5\lambda/4$，所以这两路的信号相位相加。从 A 到 C 时，顺时针与逆时针的信号分别变化了 $\lambda/2$ 和 λ，因而它们相互抵消，C 抽头处没有信号输出。最后，在 D 抽头处，两个方向的信号都变化了 $3\lambda/4$，再一次相位叠加。注意到 B 和 D 抽头处产生的信号相位相差 $\lambda/2$。这样，在 A 抽头的单端输入变成了 B 和 D 抽头的差分信号。由互逆可知，B 和 D 抽头的差分输入会在 A 抽头处产生一个单端输出。

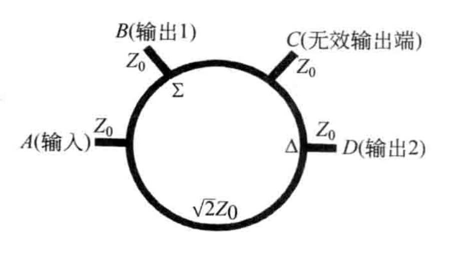


图 7.33　环形（或称鼠奔）混合型变换器

这种环形鼠奔（rat-race）混合型变换器的使用十分多样化，因为根据不同的端口输入输出指定情况可以完成不同的功能。例如，加在 B 端的信号在 A 和 C 端分成两个同相的信号（由对称性可知，加在 C 端的信号也会在 B 和 D 端产生同相输出信号）。这时，端口 D 是隔离端。反过来，在 A 和 C 端输入信号，会在 B 端相加输出。因而，B 端有时称为求和端口，以符号 $\sum$ 表示，即使是在用作输入端口时也这样标注。

加在端口 D 上的信号，会在端口 A 和 C 上产生一个差分信号输出，而 B 端现在用作隔离端口。可以推知，加在 A 和 C 端的差分信号会在 D 端产生输出。因此，D 端有时称为差分端口，以符号 Δ 表示。利用叠加原理，我们可以得出结论：加在 A 和 C 端口上的任意两个信号，在 B 端上得到它们之和，在 D 端口上则得到它们之差[32]。

重要的是要注意到，如果我们要提供阻抗匹配，那么环本身的特征阻抗必须与各抽头的阻抗相异。其精确值可以从如下的原则中推出：功率是均等地被分配到两个具有相同负载的输出抽头上。因此，从源驱动点 A 处看到的等效负载阻抗是 $Z_0/2$，而这个负载阻抗由两条并行通路驱动，每条通路的阻抗称为 Z_{ring}。我们知道，若四分之一波长传输线的特征阻抗是源阻抗和负载阻抗的几何平均，那么它可用作阻抗匹配器。这里，源阻抗是 Z_0，有效负载阻抗是 $Z_0/2$，所以满足以下条件时就可以达到匹配：

$$\frac{Z_{\text{ring}}}{2}=\sqrt{Z_0\cdot\frac{Z_0}{2}}=\frac{Z_0}{\sqrt{2}}\Rightarrow Z_{\text{ring}}=\sqrt{2}\,Z_0 \tag{50}$$

这种类型的混合型变换器称为环形（或鼠奔）混合型变换器[33]。应该清楚的是，严格的环形并不是必需的。所需的只是总周长和抽头位置满足不同的波长标准，圆环线特征阻抗是抽头阻抗的 $\sqrt{2}$ 倍。注意到，这些基于波长的标准意味着该混合型变换器必然是一种窄带元件。带宽的精确值取决于可以容忍的幅度和相位失配（或者是端口反射系数）值。尽管这么说，典型的有用带宽对于相对平坦的幅度与相位而言通常是在 15% 这个数量级（我们再次强调，精确的带宽完全是“相对平坦”指的是什么函数）。如果反射系数比幅度平坦更重要，那么带宽要大得多（如 50%）。

---

[32] 当然，这里的“任意”并不是真的任意。我们受限于混合型变换器可以正常工作的窄带宽内。

[33] 有时也称之为混合环 hybrid ring。但这里“hybrid”是名词，“ring”是形容词，所以“ring hybrid”（环形混合型变换器）更合乎语法些。

当混合型变换器用于从单端、参考端接地的输入产生差分输出时(或者反之),它也被称为巴伦(balun)变换器(平衡-不平衡变换器)。Balun 常被误发音成“bail-un”。偶尔也能看到 unbal 用来指单端到差分的变换,但 balun 更常用得多。

因为经典的环形混合型变换器的直径是波长的函数(大约是 λ/2 量级),实际实现时,在特别高频或低频,会有些问题。在非常高频,环的直径会太小,以至于与馈线宽度是同数量级的。除了仅因为这些图形的拥挤而引起的画版图的挑战外,馈线的互相接近也会严重干扰混合型变换器的工作。解决这个问题的一个方法是,将环的周长增大 λ/2 的整数倍,并且注意根据维持恰当的相位关系,来重新调整端口的位置。

在很低频时,问题就变成了尺寸过大。这种情况下,通常考虑采用集总方法实现,因为微带线混合型变换器所占的面积可能会非常大。图 7.34 所示的网络实际上是一个二分器(即,一个把信号分到两个频带上的网络——此说法不确切,译者注)。低通和高通滤波器分别产生了相位值随频率变化的滞后和超前。低通滤波器是 λ/4 传输线的集总近似,而高通滤波器是其对偶。每个滤波器的拐角频率是这样来确定的,即在中心工作频率处都产生 90°的相移。即使随频率的变化,每个滤波器的相位偏移都有变化,但在一个很宽的频率范围内,输出相位的差恒为 180°(虽然幅度仅在一个频率上严格相等)。集总法也有微带线同样的缺点,即相对来说是窄带工作。虽然缺少隔离输出端口,依然常称其为混合型变换器[34]。

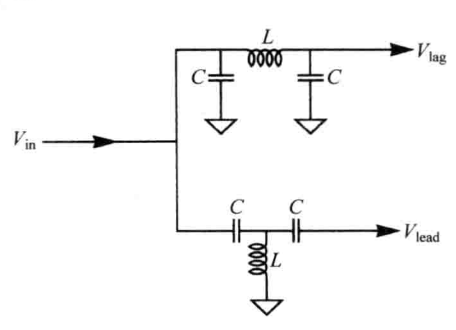


图 7.34　简单地采用集总元件的 180°分配器/合成器(二分器版本)

如果 Z_0 是驱动输入端的源阻抗,那么我们希望每个滤波器的输入阻抗都等于 $2Z_0$。如果用一个阻抗是 $2Z_L$ 的负载把两个输出端连接起来(等价于每个输出端都用一个 Z_L 接地),那么我们要使每个滤波器的特征阻抗都等于 $2Z_0$ 和 Z_L 的几何平均:

$$\left(\sqrt{L/C} = \sqrt{2Z_0Z_L}\right) \Rightarrow L/C = 2Z_0Z_L \tag{51}$$

需要再有一个方程来完成设计。这个方程可以从每个滤波器必须选择这样一个拐角频率,以使得在混合型变换器工作的中心频率点产生要求的正交相移:

$$\omega_0 = 1/\sqrt{LC} \tag{52}$$

这样就可以解得两个元件的值为

$$C = 1/\omega_0\sqrt{2Z_0Z_L} \tag{53}$$

$$L = \sqrt{2Z_0Z_L}/\omega_0 \tag{54}$$

因此,对于一个由 50Ω 驱动的、终端电阻为 100Ω(即每个输出端到地是 50Ω)的 1GHz 混合型变换器,L 和 C 分别是 11.3 nH 和 2.25 pF。分立元件或微带线元件(或是这两种的结合)都可用于实现这种混合型变换器。

经常会遇到负载电阻并联着一些寄生电容,如果需要的话,可以用恰当的电感将其谐振

---

[34] 偶然也采用这样的说法——因为两个彼此反相的输出端口在它们为具有相同幅度的信号驱动时能够互相被隔离,用混合型变换器这个词也没有错。

掉(从原则上讲)。串联在负载上的寄生元件也可用类似的方法除去(这也只是在一个窄频带上工作)。

如果需要更好地用集总元件来模拟分布式的耦合器，可以用低通 π 形网络取代单个的 λ/4 传输线段。为了再改善一下带宽，λ/2 传输线段最好用高通 T 形网络[35]。其原因同样是：相移的行为与频率成互补关系，导致了整体上是更为恒定的相移，尽管幅度随频率变化不能保持恒定。整个耦合器然后就如图 7.35 所示。

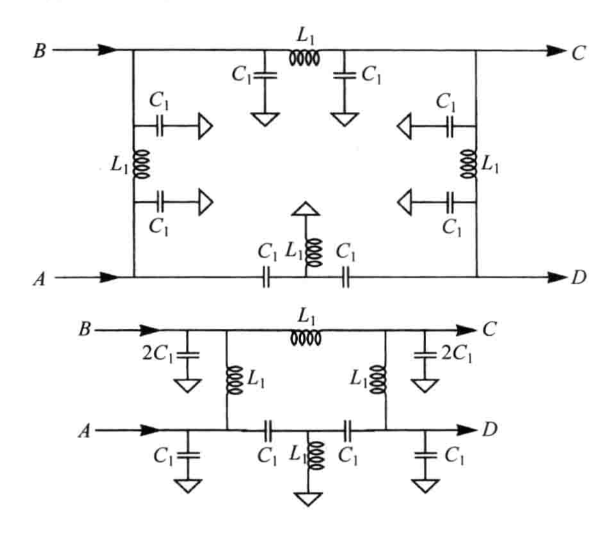


图 7.35　180°混合型变换器的一种可能的集总元件实现

元件值选取如下，在中心工作频率，所有的电抗都等于环线阻抗

$$C_1 = \frac{1}{\sqrt{2}\,\omega_0 Z_0} \tag{55}$$

$$L_1 = \frac{\sqrt{2}\,Z_0}{\omega_0} \tag{56}$$

集总方法有着与分布式方法同样的窄带宽(10%～15%)局限。集总法在较小的空间内实现比经典的分布式法有优势，因此日益受欢迎，尤其在低频率下。因为集总电路易于调整，即使对更高的频率上，集总法在某些场合下也相当有吸引力。

有些其他应用中，需要产生(或合成)两个正交信号，而不是反相信号。这样的应用包括移相器(因为不同比例的同相和正交信号合成可以得到任意的相移)、单边带生成器和检测器，以及正交调制(包括用正交信号来驱动二极管混频器来实现这种调制的重要应用)。用来实现这些功能的一个常用微带元件是 90°(正交)混合型变换器，也称为分支线耦合器或混合变换器，见图 7.36。

中心方框的每个边长都是 λ/4，所以不难看出在邻近端口上的信号是正交的。然而，除了观察到这个外，很难再通过目测来得到支路阻抗了(也就是说，除非你是事先知道答案的)。我们又一次看到，根据图 7.37 的奇偶模分析，可以简单地完成设计。为了最大限度地

[35] S. Parisi, "A lumped-Element Rat-Race Coupler"(一个集总元件的鼠奔耦合器), Applied Microwaves, Augest/September 1989, pp. 84-93.

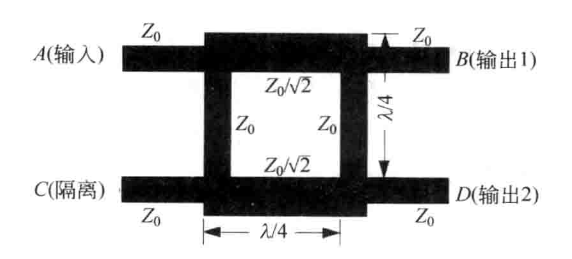


图 7.36　分支线混合型转换器(所示的是拐角未经斜角化的 3dB 版本)

利用对 Wilkinson 分配器分析的熟悉,让我们在 A 和 C 这两个相邻端口上进行奇偶模分析。若用电流源驱动,那么奇偶激励的叠加,使我们能够分析仅在主输入端加驱动时的情况。

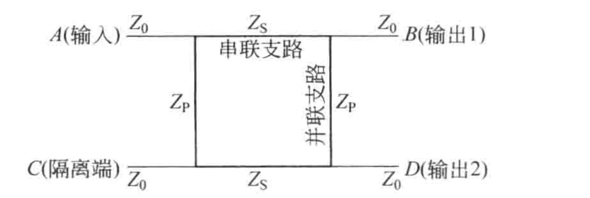


图 7.37　用于进行偶、奇模式分析的形式化分支线混合型转换器(串联臂,并联臂)

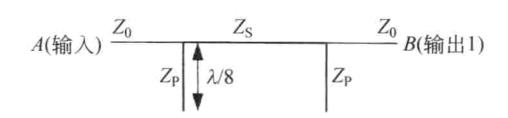


图 7.38　用于偶次模分析的等效半电路

共模驱动时电路可以对半分开分析,恰如 Wilkinson 情形,如图 7.38 所示。因为共模激励保证使上下两半电路间没有电流,对半分开得到了两个开路的线段,每个长 $\lambda/8$。每个线段都相当于导纳为 jG_P 的并联电容。采用图 7.38 的电路直接可以证明,如果当输出端接 Z_0 时,要使输入阻抗等于 Z_0,需要满足:

$$G_P^2 - G_S^2 = Y_0^2 \tag{57}$$

其中,$G_S = 1/Z_S$。通过观察奇模的相关方程,可以看到奇偶模激励适用同样的匹配约束条件。一个分量匹配了,另一个分量自然就匹配了。

可以看到,有无数多种支路阻抗组合都能达到匹配。所以我们还需要一个约束条件来确定各个特征阻抗的值。由奇偶模响应的叠加,可计算输送到输出端口的功率之比:

$$\frac{P_{out1}}{P_{out2}} = \left(\frac{Z_0}{Z_P}\right)^2 \tag{58}$$

如果要等功率分配(实际也几乎总是如此),那么解相关的阻抗得到

$$Z_P = Z_0 \tag{59}$$

$$Z_S = Z_0/\sqrt{2} \tag{60}$$

这些就是图 7.36 中所示的阻抗值。这个等功率分配经常通过称耦合器为 3dB 正交混合型变换器得以强调。

虽然在所画的分支线耦合器草图中有些尖的拐角,通常将 $Z_0/\sqrt{2}$ 那些臂的外角进行圆角化。也可以用圆形版图来代替方形版图,虽然方形的要常见得多。

就像鼠奔变换器,经典分支线混合型变换器相对来说是个窄带的器件,随频率变化,相移也会变化。有用带宽取决于对“有用”的定义,但作为一个一般的规律,分数带宽(即带宽与频带中心频率之比,译者注)是 10%～15%,这就像环形的混合型变换器(基于同样的原

因)。这个相当小的值要求有高的模型和制造精度。一个 4.7GHz 分支线耦合器的仿真(采用软件 Sonnet Lite 9.51)也很好地证明了这样的要求。虽然这个专门的耦合器设计使用了未经倒角的拐角(如图 7.36),因此不是最优的,但其呈现出的性能局限是典型的。如图 7.39 所示,当频率偏离超过设计中心频率约 5%时,耦合幅度会大大偏离 3dB 的目标值。同时,回波损耗和隔离度都迅速恶化。

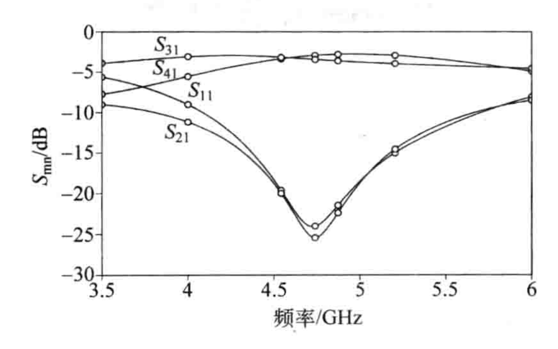


图 7.39　一个具有代表性的 4.7GHz 分支耦合器的模拟得到的性能

仿真得到的响应中的不对称性主要源自缺乏对不连续性的修正。如前所述,尖角可能产生非理想的行为。外转角增加了并联电容(至一阶效应),内转角导致耦合器的串并联支路呈现出不相等的电长度。如我们多次见到的,倒角可以修正前一种效应。对后一种效应的一阶修正可以通过将并联支路增长 d_2 实现,该值可由我们已经给出的方程推导出来(式(30)),为方便起见,这里再给出来[36]:

$$\frac{d_2}{H}=\frac{120\pi}{Z_{\text{series}}\sqrt{\varepsilon_{\text{re,ser}}}}\left[0.5-0.16\frac{Z_{\text{series}}}{Z_{\text{shunt}}}\left(1-2\ln\left[\frac{Z_{\text{series}}}{Z_{\text{shunt}}}\right]\right)\right]\tag{61}$$

所需长度对于 FR4 的分支线耦合器典型值是约 1.3H。

恰如之前的那些例子中,可以用集总法实现微带电路,只需将 λ/4 传输线段进行恰当的近似,如图 7.40 的上半部分结构。在将元件合并之后,整个集总网络包括四个电容和四个电感,如图 7.40 下方电路图所示。元件值由以下为我们所熟悉的关系式给出:

$$L_1=\frac{Z_0}{\sqrt{2}\,\omega_0}\tag{62}$$

$$C_1=\frac{\sqrt{2}}{\omega_0 Z_0}\tag{63}$$

$$L_2=\frac{Z_0}{\omega_0}\tag{64}$$

$$C_2=\frac{1}{\omega_0 Z_0}\tag{65}$$

其他的集总法也可以实现,有些还有更好的特性。例如,可以交换高通和低通部分来改善带宽,利用两个部分对随频率变化的相移的互补性,依此来提供相移对频率的一阶不敏感性[37]。

[36] 见脚注 20。

[37] 见参考文献 K. Ali and A. Podell, "A Wide-band GaAs Monolithic Spiral Quadrature Hybrid and Its Circuit Applications", *IEEE J. Solid-State Circuits*, v. 26, no. 10, October 1991.

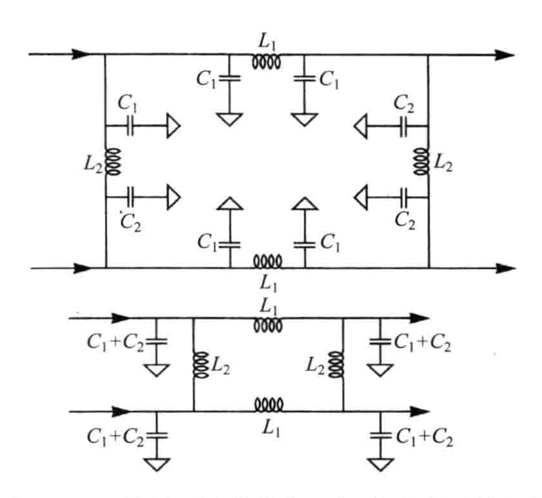


图 7.40 90°混合型变换器的一种可能集总元件实现

这种性质与之前的环形混合变换器的简单集总近似中所利用的性质相同。

后面我们会看到，在正交耦合器的诸多应用中，一个极有价值的应用是缓解功率放大器中阻抗匹配不佳的问题(见第 20 章)。

7.6.4 定向耦合器

定向耦合器，可以是混合型变换器，因其将输入功率分流到不同目的地，是一种功率分配器的子集。定向耦合器的典型结构是一个四端口器件，可以将功率以预先设定的比例进行分配，该比例通常与 1∶1 大不相同。定向耦合器的一个常见应用是信号采样器件。例如，来自功率放大器输出端的反馈信号一般由定向耦合器提供。该反馈信号可以简单用作输出功率指示，或者可形成一个用于减小失真的负反馈回路。

定向耦合器很有用的原因有几条。其一是它们可以从信号中区分出单独的正向和反向信号分量。正是因为后面这个属性，定向耦合器才得名。现代网络分析仪十分依赖于这个能力来精确测量，比如阻抗。在第 8 章我们会更全面地展开对这个课题的讨论。

另一个属性是当用作混合型变换器时，会有一个隔离端口。加在输入端口上的功率在理想情况下完全不会耦合到隔离端口。当这个隔离端口用作输入端时，隔离减小了驱动两个端口的信号源之间的相互作用。例如，在接收机上进行双音互调测试时，这个隔离就非常有用，因为如果两个信号发生器间隔离不好，就可能使这两个发生器之一工作不正常，从而导致测量失效。

如同我们看到过的其他许多例子，定向耦合器可用集总或分布式方法实现。

虽然从根本上讲定向型耦合器是个四端(或更多端)器件，它通常用作三端口元件。在所有情况下，包括未使用的端口(一般是隔离输出端)在内的所有端口，都要以其特征阻抗进行终断。大部分商用三端耦合器都有一个内部宽带终断器。

定向耦合器的品质因数包括耦合因子、隔离度和定向性。耦合因子定义为输入功率与分配到耦合端口(辅助端)的功率之比：

$$C_{\mathrm{F}} \stackrel{\text{def}}{=} \left.\frac{P_{\mathrm{IN}}}{P_{\mathrm{AUX}}}\right|_{\text{forward}} \tag{66}$$

大部分出处就用 C 表示耦合系数，但为了避免与电容的符号混淆，这里加了下标 F。但我们也笼统地用 C_F 来同时表示功率比及其分贝值。留给读者根据具体问题来判定具体是指哪个值。

耦合因子的典型值可以从 3～20dB。更大的耦合因子意味着有更多的功率耦合到(直通)主输出端，而不是辅助(耦合)输出端。术语可能有点混乱，因为高的耦合系数相当于更小的功率耦合到了耦合输出端[8]。同样地，耦合因子越低，从输入端到输出端的衰减就越大。更添乱的是，一些数据手册和教材对耦合因子有更严格的定义，为式(66)的倒数，所以用分贝表示就会有负号。虽然这种定义更正确(因为分数的功率耦合到输出端口的确是小于 1)，一般应用中，习惯上还是不带负号。

如果定向耦合器处于反向工作状态，在输出端加功率，在输入端终断，理想情况下在辅助输出端测不到信号(这就是“定向”这个术语的由来)。但是不可避免地，有些反向功率会从辅助输出端泄漏出来。描述反向泄漏被抑制的程度的一个度量就是隔离因子，定义为

$$I \overset{\text{def}}{=} \left. \frac{P_{\text{IN}}}{P_{\text{AUX}}} \right|_{\text{reverse}} \tag{67}$$

隔离因子在 30～60dB 对于商用器件是常见的。

通常将这两个量结合在一起产生一个称为定向性的品质因数 D：

$$D \overset{\text{def}}{=} \frac{I}{C_F} = \frac{P_{\text{AUX}}|_{\text{forward}}}{P_{\text{AUX}}|_{\text{reverse}}} \tag{68}$$

因此，定向性是描述耦合器区分正向和反向信号能力的量。我们希望定向性无限大，但实际所有的耦合器达不到这个理想值。微带耦合器尤其具有相对较低的定向性(例如，通常范围是 10～15dB)，这是由于奇偶模相速度的大不相同，就如我们很快会解释的。因此，在商用耦合器中，带状线(或同轴线)结构比微带线更多被采用。

为完整性起见，注意耦合器的插入损耗取决于耦合的程度。传递到耦合输出端的功率越大，剩下能输送到直通输出端的功率就越少。

假设是无损耗的耦合器，插入损耗与耦合因子的关系如下：

$$(P_{\text{out}} = P_{\text{in}} - P_{\text{coupled}}) \Rightarrow \frac{P_{\text{out}}}{P_{\text{in}}} = 1 - \frac{P_{\text{coupled}}}{P_{\text{in}}} = 1 - 10^{-C_F/10} \tag{69}$$

其中，耦合因子 C_F 是用分贝表示的。表 7.3 提供了一些数字来说明插入损耗和耦合的关系。对于耦合因子超过 10dB 或 15dB 的情况，大部分实际耦合器的插入损耗都受寄生机制控制(例如，趋肤效应或电介质损耗)，而不是受输送到耦合端口的功率控制。

表 7.3　理想的插入损耗与耦合因子的关系

耦合因子/dB	插入损耗/dB	耦合因子/dB	插入损耗/dB
3	3	15	0.14
6	1.3	20	0.044
10	0.46		

定向耦合器可以用分布式或集总方法实现。对于后一类，一个广泛应用的宽带实现是基于变压器耦合的，如图 7.41 所示。从图中可以看出，该耦合器提供了端口定义的选择(以

[8] 这就是去掉分贝表示的负号后造成的假象。

括号区分)。正如我们即将看到的,一种选择的耦合输出信号与输入信号是反相的,其他选择的耦合输出信号则与输入电压同相。

由 Sontheimer 和 Frederick 在 1969 年前后提出的诸多选择中,这个特定的选择一直都很受欢迎,因为它性能好,构造简单[39]。通常在大约两个频程(一个频程为十倍,译者注)里可以获得好的定向性。通过增加一些元件来改善频带边缘的性能,并且对剩下的缺陷利用校准来进行表征和算法补偿,这种耦合器可以工作在几个频程上。

取决于端口的设定,这个耦合器可提供与主输入同相或异相的耦合输出。图 7.41 中,在括号中给出了对应于同相版本的端口设定。我们先来分析同相的情况。由于精确的分析会使工作原理的讲述变得模糊,我们只进行简化了的推导。最终会给出完整推导的结果,包括更为准确的设计方程。

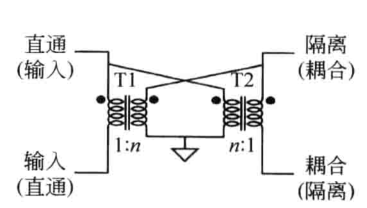


图 7.41　宽带定向耦合器(变压器版本)

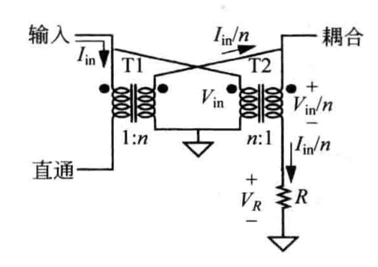


图 7.42　宽带 0°定向耦合器(近似分析)

图 7.42 中,假设耦合因子足够大(即变压器匝数比足够高),从而可以忽略 T2 中的 n 圈线圈中的电流。根据这个假设,可以把这个耦合器看作用一个变压器(T1)去感应流到输入端的电流,另一个变压器(T2)感应输入电压。前者是近似正确的,而后者则没有近似。若可以忽略流到连接在耦合端上的负载(图中没有画出)的任何电流,那么从 T2 的次级线圈和电阻端间会有与输入电流成比例的电流流经。因而,电阻上的电压与该电流成正比。如果令输入端阻抗等于 R,那么有

$$I_{in} = \frac{V_{in}}{R} \tag{70}$$

图中电阻上的电压降为

$$V_R = \frac{I_{in}}{n}R = \frac{V_{in}}{nR}R = \frac{V_{in}}{n} \tag{71}$$

T2 次级线圈上的电压与输入电压成比例(具体是 V_{in}/n)。因此,耦合输出端的电压为

$$V_{coupled} = V_R + \frac{V_{in}}{n} = 2\frac{V_{in}}{n} \tag{72}$$

注意,如果 I_{in} 符号为负(对应于反向电流,即从直通端流到输入端),那么电阻上的电压降也是负值,从而导致了在耦合端电压的抵消。这样可以看出,耦合端的输出只是对正向电流分量作出反应,而对反向分量则没有反应。

变压器的匝数比 n,决定了耦合因子。n 的值越大,馈送到耦合端的功率就越少。当耦

[39] 见 Carl G. Sontheimer and Raymond E. Frederick,"Broadband Directional Coupler"(宽带定向耦合器),U. S. Patent #3426298,1969 年 2 月 4 日授予。

合端口以电阻值为 R 的电阻端接时，电压降至式(72)给出的开路值的一半。耦合因子(dB)由下式给出：

$$C_F \approx 20\log n \tag{73}$$

进行详尽的分析并不困难，但很繁琐。由于再提供直观理解的意义不大，我们省略了分析细节，仅将结果以如下的方程组形式给出：

$$V_{\text{thru}} = V_{\text{in}}\left(\frac{2n^2-1}{2n^2+1}\right) \tag{74}$$

$$V_{\text{coupled}} = V_{\text{in}}\left(\frac{2n}{2n^2+1}\right) \tag{75}$$

$$V_{\text{iso}} = -V_{\text{in}}\left(\frac{1}{2n^3+n}\right) \tag{76}$$

我们看到，随匝数比增加，直通输出端的电压收敛到输入电压。同时，耦合输出端的电压如预期逼近 V_{in}/n。最后，隔离端的信号随 n 增大而迅速下降。也要注意在隔离端相位的反转。

另一种端口设定能使耦合器具有一个隔离端，并且能在输入端和耦合端之间产生 180°相位反转(见图 7.43)。在这种实现中，流进输入端的电流由 T1 精确采样。若可假设耦合因子是大的，那么在直通端和输入端的电压几乎相等。这样，输入电压的一个很好的近似出现在 T2 的有 n 圈的初级线圈上，与该近似电压成比例的电压则出现在 T2 的次级线圈上。假设流到连接在隔离端口上的阻抗的电流可以忽略，那么流经电阻 R 的电流为 I_{in}/n。因而，

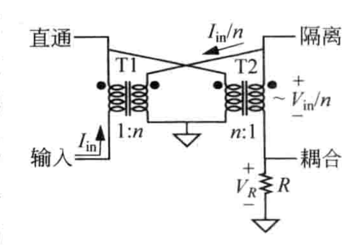


图 7.43 宽带 180°定向耦合器(近似分析)

$$V_R = -\frac{I_{\text{in}}}{n}R = -\frac{V_{\text{in}}}{nR}R = -\frac{V_{\text{in}}}{n} \tag{77}$$

其中，我们再一次假设输入端阻抗等于 R，因而有

$$I_{\text{in}} = \frac{V_{\text{in}}}{R} \tag{78}$$

注意，电压 V_R 与 T2 次级线圈两端的电压幅度相等，但极性相反。这样，这两个电压相互抵消地叠加，致使隔离端口输出为零。

耦合因子(还是以 dB 为单位)近似为

$$C_F \approx 20\log n \tag{79}$$

反向电流则在隔离端产生相长的干涉，从而可以区分出输入电流的正向和反向分量。

通过一个精确的分析，得到以下这些直通、耦合以及隔离端口输出的表达式：

$$V_{\text{thru}} = V_{\text{in}}\left(\frac{4n^4-2n^2}{4n^4+1}\right) \tag{80}$$

$$V_{\text{coupled}} = -V_{\text{in}}\left(\frac{4n^3}{4n^4+1}\right) \tag{81}$$

$$V_{\text{iso}} = -V_{\text{in}}\left(\frac{2n}{4n^4+1}\right) \tag{82}$$

从这些表达式中，我们可以看出前面所作的假设是合理的。相互抵消的不完美性导致了隔离端口的输出非绝对为零，但随匝数比增加，该输出迅速减小到一个非常小的值。类似

地，当 n 增大时，耦合端的电压幅度足够快地收敛到 V_{in}/n，所以几乎可以在所有实际情况中对于耦合因子都可使用该假设(式(79))。

除去宽带性质之外，基于变压器的耦合器适用于宽范围的端接阻抗[40]。因为变压器本身不对可使用的阻抗有什么根本的限制，同一个耦合器可以用于，比如 50Ω 或 70Ω 的系统。宽带变压器耦合器通常是缠绕在双目芯上的(即有两个相邻圆柱孔的一块磁性材料)，或在两个平行环形铁氧体芯上，或者仅是一对环形铁氧体芯。

如果宽带工作不是必须的，可以用其他替代方案来实现，从而避免许多工程师认为的采用线圈变压器的不便利性。其中一种方法用了一对集总正交移相器(在图 7.44 中使用了 $\lambda/4$ 传输线的 T 形网络近似)，它们间用电容耦合起来。

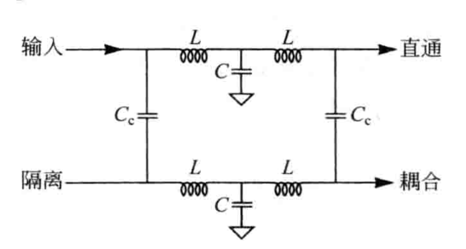


图 7.44　集总元件实现的窄带定向耦合器

为了理解这种耦合器是如何工作的，在输入端加一个测试信号(通过一个阻值为 Z_0 的电阻)。忽略左边的耦合电容 C_C，大部分输入功率直接加到直通端，在此过程中经历了正交延迟。穿越顶部延迟线功率的剩余部分向下流经右边的电容 C_C。该功率的一半分配给了耦合输出端负载，其他的则向左流经底部正交延迟线，流入隔离端。注意到这个信号经历了两次相对于输入端的正交相移，因此与输入信号的相差是 180°。我们称为反转信号。

经左边的耦合电容，从输入端到隔离端也有条直接的通路。流经该通路的信号与反转信号的幅度相等但相位精确反相，因此，它们在隔离端叠加时就会相互抵消，这也解释了该端口得名的原因。

最后，在输入端和耦合端之间还有一条通路。正向信号可由输入端进入，向下经左边耦合电容到底部正交延迟线，最后到达耦合端。该信号与经顶部延迟线，向下经右边耦合电容到达耦合端的信号同相相加(两个信号都经历了一次正交延迟和穿越一个耦合电容)。这样，输出端为受 C_C 值控制的非零值，而隔离端在理想情况下没有输出。

请注意，以上分析都隐含地假设所有端口都有合适的端接(否则，功率分配就不会像我们所描述的那样了)。

如果假设在感兴趣的带宽内，耦合电容的电抗比 Z_0 大得多(意味着我们限制耦合因子不小于，比如说 15～20dB)，这样我们能提供一个简化设计步骤，其中 T 段延迟线的元件值由熟悉的关系式给出。至少这个步骤给出了有用的初值，基于这个初值最终的设计在进行几次迭代后就能得到。在大多数情况下，该步骤本身就产生了完全实用的设计。

再次根据期望的特征阻抗来选择 L/C 之比如下：

$$\sqrt{L/C} = Z_0 \tag{83}$$

然后用 LC 之积作为期望中心频率的函数：

$$1/\sqrt{LC} = \omega_0 \tag{84}$$

这样选择元件值如下，使其电抗等于特征阻抗：

---

[40] 实际上，这两个属性在这里是相互关联的。

$$C = 1/\omega_0 Z_0 \tag{85}$$

$$L = Z_0/\omega_0 \tag{86}$$

为了完成设计，选择电容 C_C 达到所需的耦合水平。耦合到直通端口的功率比就是与 C_C 和 Z_0 有关的电压分压因子的平方：

$$\frac{(\omega Z_0 C_C)^2}{1+(\omega Z_0 C_C)^2} = 10^{-C_F/10} \tag{87}$$

这里的 C_F 是耦合因子，以分贝表示。

在耦合非常弱的情况下，我们认为与 1 相比，$(\omega Z_0 C_C)^2$ 是可以忽略的。在那种情况下，通过解公式(87)得到的耦合电容有下面的值：

$$C_C \approx (1/\omega Z_0)[10^{-C_F/20}] \tag{88}$$

用公式(85)来将耦合电容与延迟线电容联系起来也是直截了当的：

$$C_C \approx (1/\omega Z_0)[10^{-C_F/20}] = C[10^{-C_F/20}] \tag{89}$$

根据这些公式，一个 1GHz、15dB、50Ω 的耦合器要用一些 3.18pF 的电容和 7.96nH 的电感来综合这些传输线，连同接近 0.565pF 的耦合电容。又一次，后面这个值实际上比(假象地)报告有三个有效数字精度的结果更为近似一些。但是它可以作为随后的基于准确模拟或实验数据的结果改进的一个好的起点。对于这个特定的例子，仿真揭示了耦合因子极其接近设计目标：在 1GHz 时大约是 14.9dB(970MHz 是 15dB)。在 1GHz 时，被隔离的输出相对于主输入提供了 30dB 的衰减(970MHz 时是 33dB)。因此在设计的中心频率处，定向性稍微比 15dB 大一点，而在 970MHz 处则上升到 18dB。或许产生最大定向性的频率比设计的目标频率低，这一点并不特别令人惊讶，因为耦合电容增加了一些电容负载(另一负面影响是特征阻抗略微减少了)。如果绝对需要的话，可以通过减小线路电容来补充这个影响。然而，这样小的需要的修正通常在元件的标准公差之内，因此在实践中这样细微的调整是不必要的。最后，在这个特定的设计中，输入回路损耗直到 1.24GHz 都超出了 15dB。

这个耦合器的另一种实现方式是把真正的 λ/4 传输线与集总的耦合电容结合在一起，见图 7.45。这种结合的实现有时为人们所喜好，因为它消除了一大堆无源元件，而同时保持了通过适当地改变耦合电容值来改变耦合程度的便利性。当频率足够高到在合理的小的空间范围内能使传输线得到实现，这种结构是最实用的。为了对电路的性能有良好的预见性，这两个传输线需要彼此分离得足够开，来保证所有的耦合是由集总电容来提供的。通常，“充分的分开”相当于线到线之间的空间至少是电介质厚度的若干倍(例如，5 倍)。在实际中，这种分开程度经常受限于可得到的耦合电容的物理尺寸。耦合电容的值用同一个公式来计算，这个公式适用于纯集总元件的实现(方程(88))。电容的寄生参数必须被保持至足够小以至于可以被忽略，不然的话，在仿真中必须另外显式地加以包括，就与微波电路的其他分析一样。在这些限制范围内，这种特定的结构能提供极佳的性能，因为它对 FR4 的许多严重缺陷最不敏感。

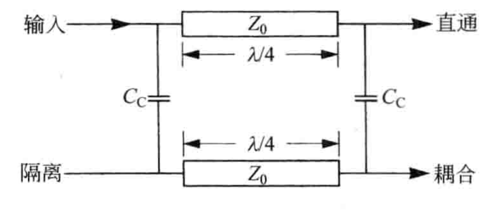


图 7.45　带有集总电容耦合的微带线实现的定向耦合器

前面这个耦合器的设计源于用分布的元件代替一些集总元件。我们还可以继续这个过

程来设计完全不用集总元件的耦合器。所需要的耦合是通过适当选择两个微带线之间的间距来提供的，如图 7.46 所示。

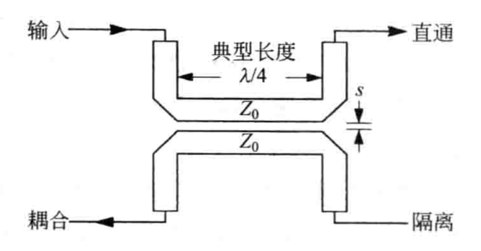


图 7.46　经典的微带线定向耦合器

初看起来，这个图中的耦合器好像是图 7.45 中的耦合器的一个微拓展，是通过横向的邻近来提供必要的耦合电容。然而，请注意在这个图中，隔离和耦合的端口已经交换了位置。为了明白为什么这不是一个标记错误，我们要认识到这两个线之间的耦合现在可能是由电场和磁场的共同作用引起的。因此，需要没有基于先前例子的先入为主的观念来分析这个耦合器是如何工作的。为此，必须首先扩展在以前引入的偶次和奇次模激励的概念。

一个孤立的微带线的特征阻抗概念是很直截了当的。对于这个量的(近似)公式是同样简单与熟悉的：

$$Z_0 = \sqrt{L/C} \tag{90}$$

其中，L 和 C 是单位长度的电感和电容值。但是，当我们将第二条线带到第一条附近时，额外的自由度就被引入了这个系统，因此我们不能用一个单一特性阻抗来表征该系统。为了去除这个泛泛说法带来的神秘感，让我们考虑一下当用共模和差模信号激励一对传输线时会发生什么。为了简化起见，让我们来考察一下耦合每一根传输线到地(及相互之间)的那些电容，如图 7.47 所示的横截面图。

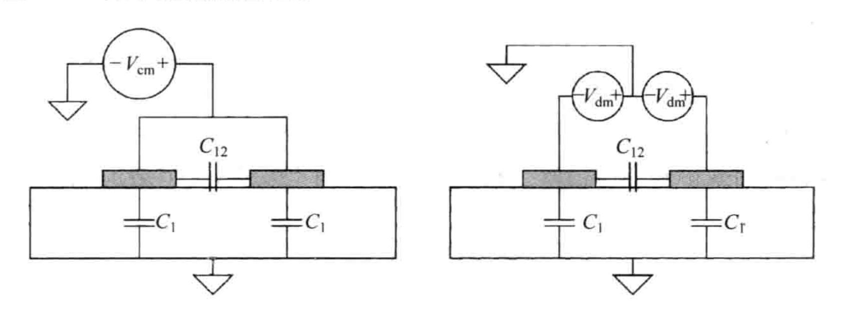


图 7.47　偶次与奇次模阻抗的确定(没有按比例画)

左侧图所示的是偶次模，相同的线电压保证了线间的耦合电容 C_{12} 不起作用(即无关)。因此，共模电容(每根线)小于单一孤立的线(其特征阻抗为 Z_0)的电容。按照惯例，我们定义偶模阻抗 Z_{0e} 等于两倍的共模阻抗。这样，当耦合强度趋近于零时，Z_{0e} 接近 Z_0。注意，Z_{0e} 比 Z_0 高[41]。

在奇次模(反对称)电压激励下，耦合电容的两端不再是零电压，因此，现在会起作用。如果我们考虑这个电容是由值为 $2C_{12}$ 的两个电容串联组成，这两个电容的中点就处于接地电势。因此，每根线的有效电容变大了，为($C_1 + 2C_{12}$)。如果定义了奇次模阻抗 Z_{0o} 等于一半的差分阻抗，那么在零耦合极限时它近似为 Z_0。我们看到，Z_{0o} 比 Z_0 小。

---

[41] 随着电流在两根线相同的方向流动，电感增加了。这种增加使得偶次模阻抗提高。类似地，在奇次模激励时反向流动的电流降低了电感，使得奇次模阻抗减少。

从以上的分析中可以明显看出，偶次和奇次模阻抗之间的差别随着传输线被置得越近而不断变大。也就是说，线之间耦合越紧，偶次和奇次模阻抗之间的差别越大。由于这种关系，这些阻抗传递与耦合因子同样的信息。因此，设计一个特定的耦合度，可以用模阻抗的确定来取代。

现在让我们推导端口电压之间的明确关系作为模阻抗的一个函数。为了简化下面的讨论，我们将端口加以标识并且区分开信号变量，如图 7.48。我们希望发现在这样一个特定情况下端口电压的关系，即当只有输入线是通过其终断（即阻抗匹配，译者注）为一电压源 V_{in} 所驱动，而其他端口都通过其各自的终断阻抗接地。注意，可以将耦合端口处理成是由一个零电压源（通过终止）驱动。就像对任意一对电压源，可以将这两个输入信号分解为一个共模和差模输入信号激励之和，如图 7.48 所示。这个共模值就是两个输入电压的平均值，或者说 $V_{in}/2$。类似地，总的差模值是两者之间的差，即 V_{in}，这个电压的每一半分别给了图 7.48 所示分解中的两个差分源。

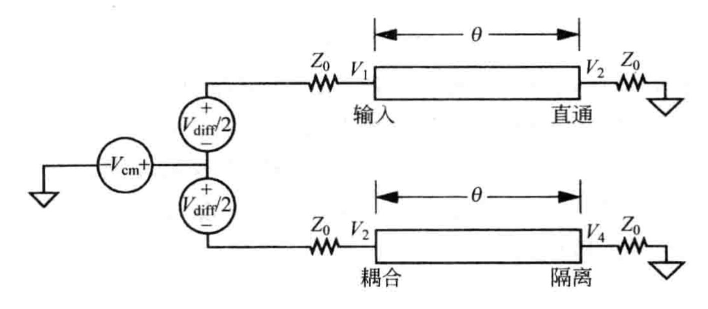


图 7.48　由通用信号源激励的端口与信号标识

为了得到相应于那个最初问题的端口电压，我们将分别计算在共模和差模激励下的节点电压值，然后相加来得到完整的解。接下来我们将马上发现，这个结构本身的对称性极大地简化这个计算过程。

首先重新给出输入阻抗为无损负载的传输线公式：

$$Z(\theta) = Z_n \frac{Z_{Ln} + j\tan\theta}{1 + jZ_{Ln}\tan\theta} \tag{91}$$

其中，Z_{Ln} 是负载阻抗，已经用 Z_n 归一化了，θ 是传输线的有效电长度（这是为了显式地兼顾 $\lambda/4$ 以外的线长的可能性，或等价地说不同于标称频率的频率情形），在这里被表示为相位角。接下来将多次用到这个公式。

首先考虑共模情况，假设差模电压为 0，因而只有共模电压激励，参见图 7.49。由于对称性，马上就可以得到 $V_{1e}=V_{3e}$，$V_{2e}=V_{4e}$。继而，应用一个电压分压器公式就得到了 V_{1e} 与 V_{1n} 的关系：

$$\begin{aligned}\frac{V_{1e}}{V_{in}/2} &= \frac{Z_{0e}\dfrac{Z_0/Z_{0e} + j\tan\theta}{1 + j(Z_0/Z_{0e})\tan\theta}}{Z_0 + Z_{0e}\dfrac{Z_0/Z_{0e} + j\tan\theta}{1 + j(Z_0/Z_{0e})\tan\theta}} \\ &= \frac{Z_0 + jZ_{0e}\tan\theta}{Z_0(1 + j(Z_0/Z_{0e})\tan\theta) + (Z_0 + jZ_{0e}\tan\theta)}\end{aligned} \tag{92}$$

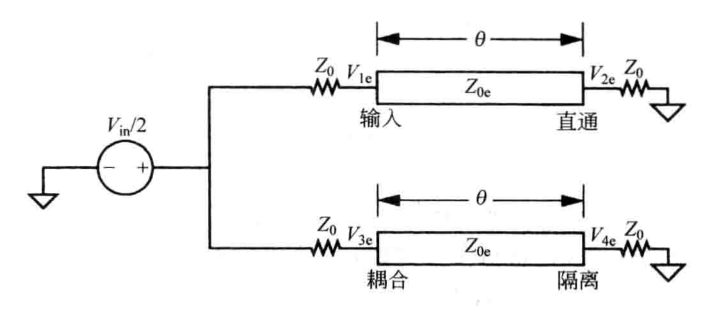


图 7.49　纯共模(偶模)激励下的耦合器

稍微简化后可写成

$$\frac{V_{1e}}{V_{in}/2} = \frac{Z_0 + jZ_{0e}\tan\theta}{2Z_0 + (j\tan\theta)(Z_0^2/Z_{0e} + Z_{0e})} \tag{93}$$

现在来考虑差模的工作情况，如图 7.50 所示，由(反)对称性可知，$V_{1o} = -V_{3o}$，$V_{2o} = -V_{4o}$。同样可以由反应器公式得到 V_{1o} 与 V_{in} 的关系：

$$\frac{V_{1o}}{V_{in}/2} = \frac{Z_0 + jZ_{0o}\tan\theta}{2Z_0 + (j\tan\theta)(Z_0^2/Z_{0o} + Z_{0o})} \tag{94}$$

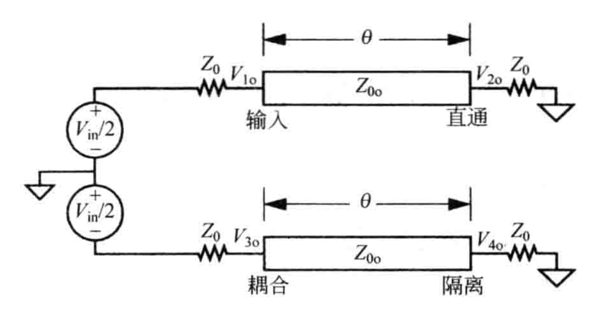


图 7.50　由纯差模(奇模)激励的耦合器

于是，通过上面的基模、偶模激励下结果相加，可以计算出传输线的输入端电压 V_1 为

$$\begin{aligned} V_1 &= V_{1e} + V_{1o} \\ &= \frac{V_{in}}{2}\left[\frac{Z_0 + jZ_{0e}\tan\theta}{2Z_0 + (j\tan\theta)(Z_0^2/Z_{0e} + Z_{0e})} + \frac{Z_0 + jZ_{0o}\tan\theta}{2Z_0 + (j\tan\theta)(Z_0^2/Z_{0o} + Z_{0o})}\right] \end{aligned} \tag{95}$$

如果主线是与信号源匹配的，那么括号里的项一定就是 1，利用这个条件，最终可以写成：

$$Z_0 = \sqrt{Z_{0e}Z_{0o}} \tag{96}$$

由于这种对称和互易性，满足公式(96)就提供了在所有四端口上的阻抗匹配(不仅仅局限于输入到主线之间)。但是必须注意到对于共模或者差模波来说，阻抗并不分别匹配。事实上，我们看到差模的阻抗要小于 Z_0，而共模的阻抗要大于 Z_0。这个不对等的事实显而易见地说明了想在单一的终断下实现两个分量的同时匹配是不可能的。当信号到达线的终端时每个分量都会有反射发生。但在满足公式(96)的情况下，这些反射恰好抵消从而实现了整体的匹配。

在耦合端口的电压也可以轻松地用共模和差模电压相加的方法得到：

$$V_3 = V_{3e} + V_{3o}$$

$$= \frac{V_{\text{in}}}{2}\left[\frac{Z_0 + \text{j}Z_{0\text{e}}\tan\theta}{2Z_0 + (\text{j}\tan\theta)(Z_0^2/Z_{0\text{e}} + Z_{0\text{e}})} - \frac{Z_0 + \text{j}Z_{0\text{o}}\tan\theta}{2Z_0 + (\text{j}\tan\theta)(Z_0^2/Z_{0\text{o}} + Z_{0\text{o}})}\right] \tag{97}$$

同样通过简化可知：

$$V_3 = \frac{V_{\text{in}}}{2}\left[\frac{(\text{j}\tan\theta)(Z_{0\text{e}} - Z_{0\text{o}})}{2Z_0 + (\text{j}\tan\theta)(Z_{0\text{e}} + Z_{0\text{o}})}\right] \tag{98}$$

在这里，方括号内的项的幅度值就是（小于 1 的）耦合因子 $|V_3/V_1|$。当 θ 是 $\pi/2$ 的奇数倍，对应于长度等于 $\lambda/4$ 的奇数倍时，这个幅度值达到最大，这也就解释了为何 $\lambda/4$ 耦合器是如此被广泛应用[42]。这个耦合的最大值很容易从公式(98)得到：

$$C_{\text{F}} = \frac{Z_{0\text{e}} - Z_{0\text{o}}}{Z_{0\text{e}} + Z_{0\text{o}}} \tag{99}$$

稍做整理后得到：

$$\frac{Z_{0\text{e}}}{Z_{0\text{o}}} = \frac{1 + C_{\text{F}}}{1 - C_{\text{F}}} \tag{100}$$

将这个公式与公式(96)结合，可以得到：

$$Z_{0\text{e}} = Z_0\sqrt{\frac{1 + C_{\text{F}}}{1 - C_{\text{F}}}} \tag{101}$$

$$Z_{0\text{o}} = Z_0\sqrt{\frac{1 - C_{\text{F}}}{1 + C_{\text{F}}}} \tag{102}$$

正如前面所说的，我们因此看到归一化的共模和差模阻抗确实表达了与耦合因子同样的信息，因此它们可以交换使用。

有了这些关系，可以写出

$$V_3 = \frac{V_{\text{in}}}{2}\left[\frac{\text{j}\tan\theta\dfrac{Z_{0\text{e}} - Z_{0\text{o}}}{Z_{0\text{e}} + Z_{0\text{o}}}}{\dfrac{2Z_0}{Z_{0\text{e}} + Z_{0\text{o}}} + \text{j}\tan\theta}\right] = \frac{V_{\text{in}}}{2}\left[\frac{(\text{j}\tan\theta)C_{\text{F}}}{\dfrac{2Z_0}{Z_{0\text{e}} + Z_{0\text{o}}} + \text{j}\tan\theta}\right]$$

$$= \frac{V_{\text{in}}}{2}\left[\frac{(\text{j}\tan\theta)C_{\text{F}}}{\sqrt{1 - C_{\text{F}}^2} + \text{j}\tan\theta}\right] \tag{103}$$

由此可见，当分母中的正切项远大于另一项时，耦合因子与刚才所求得的最大值几乎没有什么不同。正是因为这个易于满足的条件，耦合因子在一个相对宽的频带内（比如，变化一倍）近似为常数，其频带中心频率满足标称的 $\lambda/4$ 条件。

有时候对一个结果进行思考会产生比答案更多的问题。因此你也许认为最好不要想得太多，当然这取决你的哲学思维倾向。现在这里就是这个情况。当然大多数对这个耦合器发表了的描述都是基于纯数学理论从而避开了这样一个难题：即在推导在耦合端口上产生的信号表达式时，我们应用了在代数上是恰好相反的共模和差模激励。由此，我们可能会预期它们叠加的结果是几乎互相抵消，因此在耦合端口上总是产生可以忽略的电压。然而实际情况并不是如此。这看起来自相矛盾，正确的解释是，正如前面所提到过的，共模信号和差模信号在到达传输线末端时会产生反射。就如与主线的情况，这两个分量经历了极性

[42] 见，比如说，B. M. Oliver, "Directional Electromagnetic Couplers"（定向电磁耦合器），*Proc. IRE*, v. 42, November1954, pp. 1686-1692.

相反(但幅度相同)的反射。然而与主线不同是,在开始时这两个激励信号间就存在一个反相关系(导致在隔离端口处的宽带破坏性的干涉)。结果是,这两次极性的翻转产生了这样的反射,它们在耦合端口处相叠加。传输线间的耦合越紧密,这两个模阻抗的差异越大,因而这些反射也就越强。因为这个耦合端口的输出就是由这些反射波的作用而产生的,这个器件有时候也叫做背向传播波(backward-wave)耦合器。我们很快就会看到为什么要花费这么多时间来啰里啰嗦地解释这个耦合器的工作原理了。

可以采取类似的方法来推知在隔离端口处电压理想值也为零。至少从原则上说,这个结果与频率无关,因为这同样是因为共模波与差模波恰好抵消。但是,就微带线而言,实际上的隔离并不完美,那是由于所要求的为破坏性的干涉而导致奇妙的抵消实际上很难实现。这其中的原因过会儿就要讨论到。

最后,再有一点附加的工作,就可以证明在直通端口处的信号是由下式给出:

$$\frac{V_2}{V_1}=\frac{\sqrt{1-C_F^2}}{\sqrt{1-C_F^2}\cos\theta+\mathrm{j}\sin\theta} \tag{104}$$

在耦合非常微弱的极端下,公式(104)可以很好地近似为

$$\frac{V_2}{V_1}=\frac{\sqrt{1-C_F^2}}{\sqrt{1-C_F^2}\cos\theta+\mathrm{j}\sin\theta}\approx\frac{1}{\cos\theta+\mathrm{j}\sin\theta}=\mathrm{e}^{-\mathrm{j}\theta} \tag{105}$$

这反映了这样一个事实:在耦合非常微弱的情况下,几乎所有的能量都直接传输到了直通端口,在这个过程中只是经受了一个 θ 相位滞后。

我们已经推导了耦合因子作为那些模阻抗和传输线的电长度的函数关系。然而,要完成一个实际的设计,我们需要把这些量与版图尺寸联系起来。令人遗憾的是,看上去并不存在有任何简单而通用的对于微带线的公式,能够给出精确的用于从希望实现的耦合得到所必需的间距值[43]。具有足够实用精度的最简单的那些公式或许就是由 Akhtarzad 等人给出的公式,遗憾的是这套公式必须采用隐式解来得到线宽和线间距[44]:

$$\frac{W_e}{H}=\frac{2}{\pi}\operatorname{arcosh}\left(\frac{2d-g+1}{g+1}\right) \tag{106}$$

$$\frac{W_o}{H}=\begin{cases}\dfrac{2}{\pi}\operatorname{arcosh}\left(\dfrac{2d-g-1}{g-1}\right)+\dfrac{4}{\pi(1+\varepsilon_r/2)}\operatorname{arcosh}\left(1+2\dfrac{W/H}{S/H}\right), & \varepsilon_r<6 \quad (107)\\[2ex] \dfrac{2}{\pi}\operatorname{arcosh}\left(\dfrac{2d-g-1}{g-1}\right)+\dfrac{1}{\pi}\operatorname{arcosh}\left(1+2\dfrac{W/H}{S/H}\right), & \varepsilon_r\geqslant 6 \quad (108)\end{cases}$$

在这些公式中,注意:

$$g=\cosh\left(\frac{\pi S}{2H}\right); \tag{109}$$

---

[43] 十分精确但复杂的解析表达式由 M. Kirschning and R. H. Jansen 在下述文章中给出:"Accurate Wide-Range Design Equations for the Frequency-Dependent Characteristics of Coupled Microstrip Line"(用于互相耦合的微带线的频率有关的特性的精确,宽适用范围的设计方程),*IEEE Trans. Microwave Theory and Tech.*, v. 32, no. 1, January 1984, pp. 83-90。这些方程给出的答案与用场解法器进行的仿真和实验测量得到的结果十分接近(在百分之几内)。这些方程包括了色散、损耗、不相等的模速及其他一些实际的效应。商用仿真器 *APLAC* 中就采用了这些方程。

[44] Sina Akhtarzad, Thomas R. Rowbotham, and Peter B. Johns, "The Design of Coupled Microstrip Lines"(耦合微带线的设计),*IEEE Trans. Microwave Theory and Tech.*, v. 23, June 1975, pp. 486-492. 也可参考 R. Garg and I. J. Bahl, "Characteristics of Coupled Microstriplines"(耦合微带线的特征),*IEEE Trans. IEEE Trans. Microwave Theory and Tech.*, v. 27, June 1979, pp. 700-705, with corrections (更正) in v. 28, March 1980, p. 272。

$$d = \cosh\left[\pi\left(\frac{W}{H} + \frac{S}{2H}\right)\right]. \tag{110}$$

比值 $W_e/H, W_o/H$ 是对那些单根微带线的，其特征阻抗分别为 $Z_{0e}/2, Z_{0o}/2$，这样就可以使用前面所提到的针对一般微带线的公式来计算。完成这一步后，再用 Akhtarzad 的公式迭代求得耦合线的实际线宽和线间距。

之前所提到的公式就是为模拟器 Puff 用来进行理想的耦合线的仿真。但这些公式不包括不同模速度的差异(以及导体和衬底的损耗的差异)。所幸的是，Puff 也能提供考虑到这些效应更完整的模拟。这样，可以利用 Akhtarzad 等人的公式来进行初始的设计，以得到某些线尺寸的初始数据，然后用更精确的仿真来评估这个设计。因为这套理想的方程具有足够的精度来产生一个合理的首次设计，一般来说仅需要反复一两次就可以得到符合要求的最终设计。

如果我们承认有迭代的必要，那么初始设计不那么精确也是可以接受的。假如这样的话，之前所提到的那些公式的复杂性就完全可以避免，如果忽略 Akhtarzad 对差模下宽度高度之比的方程中第二项，就可以得到间距与高度的比值的一个闭合近似式：

$$\frac{S}{H} \approx \frac{2}{\pi}\operatorname{arcosh}\left\{\frac{\cosh\left(\frac{\pi}{2}\frac{W_o}{H}\right) + \cosh\left(\frac{\pi}{2}\frac{W_e}{H}\right) - 2}{\cosh\left(\frac{\pi}{2}\frac{W_o}{H}\right) - \cosh\left(\frac{\pi}{2}\frac{W_e}{H}\right)}\right\} \tag{111}$$

一旦得到 S/H，接下来就可以借助公式(106)或者公式(107)来获得 W/H。当确定了耦合部分的长度后(比如 $\lambda/4$)，这个设计也就大功告成了。再一次，因为共模和差模速度的相异，传统的做法就是利用这两个模的平均速度来选择这个长度。

有一个计算 S/H 的闭合公式简化了设计步骤，利用电子数据表就可以快速完成设计。但是这个流程依然不够简单。因此我们将介绍另一种替代方法，即适合于 50Ω 系统中弱耦合(10dB 或者更大的损耗)线的准经验公式。它非常的简单方便，仅仅只需要普通手持计算器上的少数几次按键，在速算或者粗略估计时非常有用[45]：

$$\frac{S}{H} \approx 1.11\ln\frac{0.32}{1 - \sqrt{\frac{1 - C_F}{1 + C_F}}} \tag{112}$$

其中 C_F 是耦合因子，它是一个纯比值，而不是以分贝表示。对非常小的 C_F 值，这个公式可以简化为下面的近似表达式：

$$\frac{S}{H} \approx 1.11\ln\frac{0.32}{C_F} \tag{113}$$

这个公式是由于联合了差模阻抗和耦合因子之间的解析关系而得来的

$$\frac{Z_{0o}}{Z_0} = \sqrt{\frac{1 - C_F}{1 + C_F}} \tag{114}$$

结合一个归一化差模阻抗为 S/H 函数的经验公式

$$\frac{Z_{0o}}{Z_0} \approx 1 - 0.32\exp\left(-1.11\frac{S}{H}\right) \tag{115}$$

最后，我们做这样一个假设：在弱耦合下，这些线均保持着在未耦合时的微带线的宽度。实践中，这些线宽随着耦合程度的变强而减小，但是线宽在耦合弱的情况下减少不多，因此我

[45] 这个公式主要是基于 *RFSim*99 对 1.6mm FR4 的仿真结果与 *Sonnet Lite* 场计算器仿真结果的比较。

们忽略任何宽度的调整。

对非关键的设计或者仅产生一个合理的初值以用于进一步的改良，我们足够粗糙公式(112)还是很不错的。但如果耦合值大于 10dB，这个公式所代表的行为就不正确了，因为这个公式是针对在实际应用中经常遇到的边缘耦合的结构(10～30dB 的耦合因子)做了优化。表 7.4 比较了公式(112)所给出的预测值与利用仿真工具 RFSim99 和 Puff 仿真得到的值的比较。

表 7.4　经验公式与 RFSim99 的比较

(FR4，$H=1.6$mm，$\varepsilon=406$)

C_F/dB(目标)	S/mm(Puff)	W/mm(Puff)	S/mm(公式)	C_F/dB(公式)	C_F/dB(RFSim99)
10	0.21	2.49	0.24	10.9	9.2
15	0.76	2.88	1.183	13.6	13.9
20	1.70	3.02	2.15	18.2	18.9
25	3.39	3.10	3.10	26.5	25.8
30	9.73	3.19	4.14	57.5	41.7

表格的第二列和第三列给出了 Puff 计算出的线间距和线宽，而 Puff 这个程序是用了 AKhtarzad 那些公式。得到由粗糙公式(112)计算出的放在第一列的耦合因子所需要的间距在第四列给出。为了做一个对比，在最后两列总结了分别用粗糙公式(112)以及 RFSim99 计算得到的耦合因子，这些结果是对应于由 Puff 和 RFSim99 给出的线间距及由 RFSim99 给出的线宽。RFSim99 模拟器所使用的算法没有相关的文献，我们之所以列出它的结果只是为了简单地证明我们的粗糙公式相对于现有那些仿真工具来说还是很不错的。可以见到，粗糙公式的结果与 RFSim99 以及 Puff 结果在耦合因子为 10～25dB 的范围内符合得不错。所有这三种结果在 30dB 情形下差异很大。所幸的是，这么微弱的耦合因子在实际应用中很少碰到。同时可以注意到要达到 10dB 的耦合因子需要非常小的间距。对这种结构来说，要实现更紧的耦合在实际实现中是行不通的。最后，注意到 Puff 的线宽计算表明耦合因子在 15～30dB 范围变化很小。因此，用对于一根孤立的微带线计算得到的微带线宽度就足以满足需要，至少对于初始设计是如此。

如果耦合器用图示的微带线制作，这种耦合器能提供的定向性是比较差的。这是因为由于微带线的不均匀性，共模和差模速度不是一样的。为了搞清楚这种非均匀性是如何产生不相等的模速度，比较一下对共模激励与差模激励的容性边缘场。对差模来说，边缘效应远比共模强。这是因为由于存在大量的线与线之间的边缘，在衬底上方的边缘场比共模要多许多。相应的对差模的有效电介质常数也要低，因此相位速度要更高。因为这两种相位速度不一样，最多只有对一种模式而言线的长度是 $\lambda/4$。通常是计算对两种模式的平均值为线长 $\lambda/4$。模速不等意味着相移是不完美的。这又会降低隔离度，因此特别是降低定向性。正因为如此，商用的耦合器几乎从来不采用微带线制作，而是利用带状线或者同轴线结构。

如图 7.51 所示，一个 20dB 耦合器的 Puff 仿真结果强调了不相等的模速降低微带线隔离性能的程度。没有达到无穷大的隔离(因此无穷大的定向性)，在隔离端口的信号并没有比耦合端口强度低多少。这个定向性勉勉强强超出了一个可怜的 4dB，而一个典型的带状线的实现有望达到 30dB 的定向性。事实上，这个特定的微带线例子的性能是如此之差，以

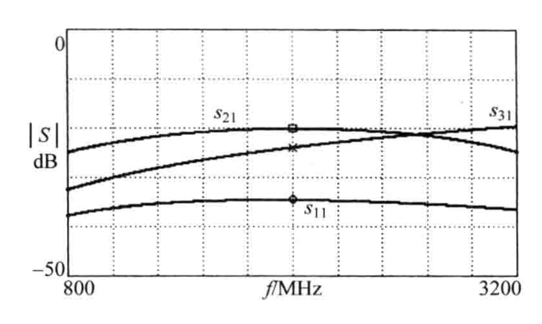


图 7.51　Puff 程序对 2GHz，20dB 耦合传输线耦合器(在 1.6mm FR4 PCB 板上实现)的模拟结果

至于在隔离端口的输出功率实际上在高于 2.6GHz 时大于耦合端口的输出功率。一个伴随的想象是在隔离端口与输入端口的信号之间的相差不是正交(即不是相差 90°)。在标称的中心频率处,隔离端口的输出滞后输入将近 180°。而在另一方面,正如所预期的,耦合输出端口和输入端口的匹配都没有出现明显的变坏。例如,耦合带宽仍然保持大约两倍。

从前面的定性讨论,我们不仅仅只知道奇模相速是太快了,也知道其原因所在。因此,可以提出一些方法来减少速率的不一致性。在多年来提出的一些这方法中,对于频率不高的情况,最常用的方法是简单地在空隙上加某个电容。在那里增加电容(理论上)不会影响偶模速率,因此这种方法代表了实现使两种相位速率相等的一个简单方法。一种实现需要的电容的方式是用蜿蜒状的结构(比如锯齿状的边缘)来增加线到线之间的侧墙电容。这种方法很适用于紧耦合的耦合线,但是缺少简单的描述公式使得设计这种补偿型的耦合器必然是一个相当艰难的任务。采用放在一般耦合线的末端的离散型电容要容易得多。通过对称地滑动一对沿着传输线的固定值电容的位置来调节达到最大的定向性方向,比在末端采用可变电容的方法要好得多,具体见图 7.52。如果调整仔细,利用这些电容能够改善定向性高达一个典型值为 10dB(一般可使定向性提高 10dB)。然而在高频下,也许很难找到或制作出具备好的品质因数和尺寸合适的电容器。另外,只是为了避免将电容器连接到微带线时的寄生效应,都随着频率的升高变得越来越困难。

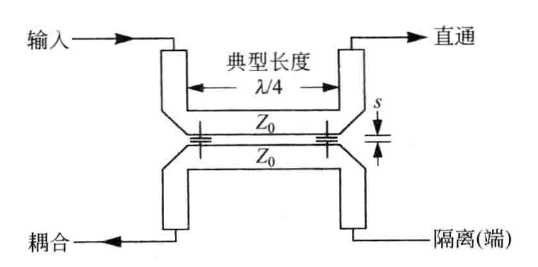


图 7.52　被加以补偿的微带线定向耦合器

在增加电容不切实际的这些频率处,可在微带线上面直接淀积电介质(这会有效地使得这种结构更像带状线)或者在这些微带线上加一层接地的屏蔽平板(再一次,使得这种结构更像带状线)。这些方法工作得不错,但是大大增加了制作的复杂度,因而抵消了微带线结构的主要优势(制作简单,译者注)。

一个完全不同的改善定向性的方法是完全避免对两种不同模的依赖性。通过消除对通过干涉来产生耦合和隔离输出端口的依赖性,具有高定向性的带宽可以大大得以扩展。

具有这些特性的一个结构是前向波或者共方向性的耦合器[46]。它表面上看起来像传统的后向波(backward-wave)耦合器,但是通过逐渐变化末端的传输线尺寸,可以抑制后向波的产生,因而可以提供一个良好的同时对共模和差分模成分的宽带终断[47]。当来自两种模式的反射都得到了压制时,就不会在离输入最近的端口上再有两个相互能增强的干涉的成分。因此,这个端口现在变成了隔离端口,而以前为隔离端口则变成了耦合端口。

尽管后向波耦合器的这些特性,共方向耦合器很少使用,这主要是因为没有简单的设计方程。因此在多数情况下,共方向耦合器仍然只是学术研究好奇心所致。不过它作为微波界新奇的结构这件事的日子最终是到了头了,因为一个简单而优美形式的共方向耦合器最近被报道出现了[48]。如图 7.53 所示,这种版本的前向波耦合器只是由两个 1/4 圆弧组成,两者间通过一个空隙隔开。圆弧半径和最小的弧到弧的空隙是两个设计的自由度,它们给出了足够窄的设计空间以允许仅借助于一个电磁场解法器就可以快速得到收敛的设计。一个利用这种方法设计的 12dB 共方向耦合器,在 50～110GHz 的二倍频频带内达到了 20dB 方向性[49]。利用标准的微带线实现的没有加补偿的后向波耦合器很难得到这种程度的性能。

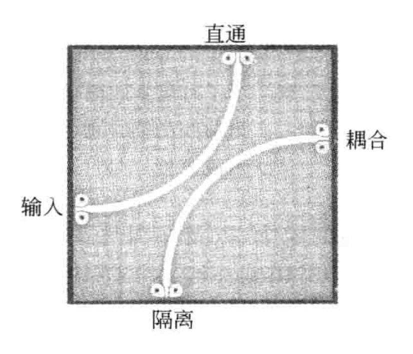


图 7.53 共向圆弧耦合器(得到 Matthew Morgan 与 Sander Weinreb 的同意)

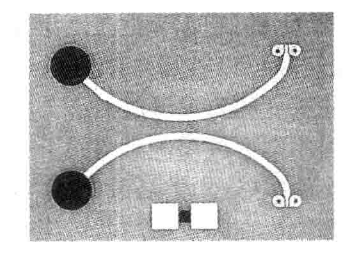


图 7.54 用圆盘终端的示意图(得到 Matthew Morgan 与 Sander Weinreb 许可)

在这些频率下做精确测量又是另外一个挑战。一个困难就是如何提供好的终断。为了闪避不可避免的由几何尺寸变化带来的性能变差,最好是在片上实现终断。然而,即使在片的终断在毫米波频段下依然会有差的性能,这是由于,比如说,与通孔相关的不连续性。为了解决这些问题,可以采用方块电阻(sheet resistance)为每方块 50Ω 的材料制成的圆盘型焊端(如图 7.54 所示)。由于这个几何形状的对称性,在圆盘上产生的任何可能的高阶模

---

[46] Pertti K. IKäläinen and George L. Matthaei, "Wide-band, Forward-Coupling Microstrip Hybrids with High Directivity"(宽带、前向耦合的具有高定向性的微带线混合耦合器), *IEEE Trans. IEEE Trans. Microwave Theory and Tech.*, v. 35, August 1987, pp. 719-25.

[47] 如果要用楔形线为偶模和奇模同时提供良好的匹配,那么这两种模的阻抗不能相差太大。因此,为了给同向耦合提供高的定向性,耦合必须相对弱一些。

[48] Matthew Morgan and Sander Weinreb, "Octave-Bandwidth High-Directivity Microstrip Codirectional Couplers"(双倍频带宽的高定向性微带线共方向性耦合器), *MTT-S Intl. Microwave Symposium*, Philadelphia, 2003. 这一类耦合器在实践中只能提供相对来说不太精确的耦合值,这是因为在紧密耦合时要抑制反向波是极其困难的。

[49] 同注解 48。

信号仅激励一整方块的材料，这样就保持电阻阻值为常数。圆形焊端的直径决定了最低端的频率，这个直径应该至少为在感兴趣的频段中的低端频率所对应的 λ/4 那么大。频率上限是由馈线中高阶模开始产生的频率决定，或是由不想要的（但总是存在的）制作缺陷引起的非对称模的激励决定。

然而，对于要求耦合非常紧密的产品，共向或普通反向波耦合器都不适用[50]。正如我们在混和正交耦合器的情形中看到的，等功率分配是经常准确要求的，但这种 3dB 耦合（等功率分配的另一种叫法，译者注），用普通的边缘耦合线是实践中做不到的，即使采用高介电常数的衬底。在某些情况下，图 7.55 所示的结构能提供多多少少是更高的耦合而不要求精密的制造技术，因此会被优先选用（对于手工原型制作而言，如果没有其他的选择）[51]。自然地，这种选择只适合在该频率范围内使用（和封装）能够得到的离散电容器是可以被容忍的情况，而且只对于耦合不是太紧的情形。

注意到一对耦合线只用到每条线的一个侧边（来进行耦合，译者注），人们会想到如果采用更多的线是否是一种替代的办法来获得更大的耦合，条件是给定了线间的间距。例如，人们会建议在多指的布局中采用两对线，如图 7.55 所示。

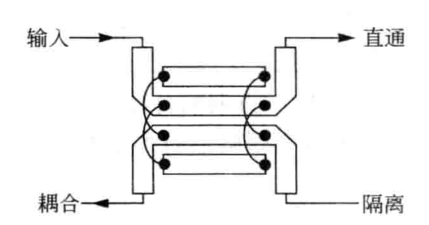


图 7.55　为了有更紧的耦合而采用了附加线的耦合器（线宽画得不成比例）

这两对耦合线具有了三对，而不是一对耦合边。从而，一个给定的线间隔能产生更加紧密的耦合。或者说，更大间隔可以被用来取得一个指定的耦合程度，这样对尺寸精度的要求就放松了。显然，这个相互交错的多指过程原则上可以一直继续下去。但是逐次消失的回报的定则，以及在图 7.55 中实际制造很狭小的线（这样做是需要的，是因为要进行平行的连接）与用键合线来实现互连的困难，使得大多数工程师仅使用两对线，尽管有时候会使用更多的线。

要解决耦合这个问题经常迫使我们来解决另一个问题。如果寻找更紧密耦合是为了在实践中实现一个 3dB 耦合器，那么保存对称性以避免降低一个 3dB 耦合器能提供的输出质量的衰退是重要的。这样一个假设不会是一个大的跳跃。遗憾的是，直通与耦合端口的输出出现在相对角的端口上，而不是在同一边。因此，由于这个根本性的缺乏输出的对称性，不可避免存在着一些性能的衰减。

在 1969 年，德州仪器公司（TI）的一个工程师 Julius Lange，描述了一个独特的方案来同时解决耦合和对称问题[52]。他的耦合器设计聪明地用将其中一条线分裂成两条的办法来增强多指交叉。用这种方法，经简单的重新排列，允许直通和耦合两个输出都在同一边出现，这样就能够提供接到这个耦合器上的电路一个对称的接口。以其最简单的形式，Lange 耦合器如图 7.56 所示。如果附加金属层是可以被利用的，那么他们可以被用来替代图中的键合线（与通孔连接一起）。在所有这些情况中，我们必须仔细来确保这些互连接的寄生参

[50] 如果有额外的导体层可被利用，那么可以一根线放置在另一根线之上，这种面面耦合（宽面，broadside，耦合）线比起侧边耦合线耦合要紧密得多。当我们研究各种实现滤波器的方法时，这个情形会重新出现。

[51] 这个说法是假设工作的频率足够的低，以至于离散元件的寄生参数可以被放心地忽略。

[52] J. Lange，"Interdigitated Stripline Quadrature Hybrid"（多指带状线正交混合型耦合器），*IEEE Trans. Microwave Theory and Tech.*，v. 20，December 1969，pp. 1150-1151.

数不会使性能降低。

正如前面所提到的，使用平行线意味着各条线本身必须具有这样的特征阻抗，其值比终断阻抗要高。因此这些耦合线比起端口的馈线要狭窄（尽管在图 7.56 中显示的是相等的相对大小）。更典型的却是，一个真实的版图看上去会如图 7.57 所示。从这个典型的版图中，读者就可以体会到所涉及的制造有多复杂（特别是，如果有更多指被使用），也可以意识到在被耦合的部分和那些更宽的、用于连接四个端口的馈线间制作一个好的过渡带来的挑战。

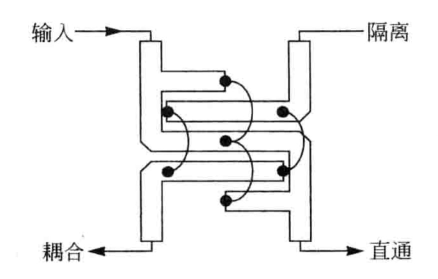


图 7.56　Lange 耦合器（线宽不成比例）

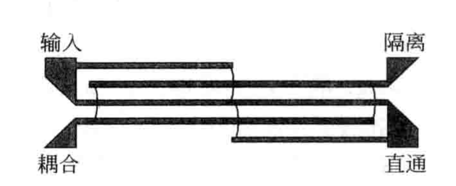


图 7.57　一个用微带线实现的 Lange 耦合器的更有代表性的版图（采用键合线的版本）

除了这些考虑之外，与 Lange 耦合器的设计涉及相对更为复杂的分析公式。即便已经有这些额外的复杂度，最终仍需借助于电磁场解法器来进行最后的细调。在挑明了这个局限性后，我们提供下面一系列的方程，可以用这些方程得到初步的第一轮设计[53]。

首先，根据作者 Ou 的文献[54]有

$$Z_0^2 = \frac{Z_{0e}Z_{0o}(Z_{0e}+Z_{0o})^2}{[Z_{0e}+(k-1)Z_{0o}][Z_{0o}+(k-1)Z_{0e}]} \tag{116}$$

和

$$C = \frac{(k-1)Z_{0e}^2-(k-1)Z_{0o}^2}{(k-1)(Z_{0e}^2+Z_{0o}^2)+2Z_{0e}Z_{0o}} \tag{117}$$

其中 k 是插指的数量。这里 C 是直接耦合因子，不是以 dB 表示的。

Osmani 然后结合 Ou 的推导给出了可实际用来设计一个 Lange 耦合器[55]的一系列方程，首先定义一个因子 q 为

$$q=[C^2+(1-C^2)(k-1)^2] \tag{118}$$

然后计算奇模阻抗

$$Z_{0o}=Z_0\sqrt{\frac{1-C}{1+C}}\,\frac{(k-1)(1+q)}{(C+q)+(k-1)(1-C)} \tag{119}$$

再计算偶模阻抗，用公式

---

[53] 这个步骤的连贯的介绍可以在下述书中找到：T. C. Edwards and M. B. Steer, *Foundations of Interconnect and Microstrip Design*（互连与微带线设计基础），3rd ed.，Wiley, New York, 2000，以及 V. Fusco, *Microwave Circuits*（微波理论与技术），Prentice-Hall, Reading, MA, 1987.

[54] W. P. Ou, "Design Equations for Interdigitated Directional Coupler"（插指式定向耦合器的设计方程），*IEEE Trans. Microwave theory and Tech.*，v. 26, October 1987, pp. 801-805.

[55] R. M. Osmani, "Synthesis of Lange Couplers"（Lange 耦合器的综合），*IEEE Trans. Microwave theory and Tech.*，v. 29, February 1981, pp. 168-170.

$$Z_{0e} = Z_{0o} \frac{C+q}{(k-1)(1-C)} \tag{120}$$

最后，用 Akhtarzad 的方法来确定实际导体的尺度和间隔，这样就完成设计。

就像支线耦合器，3dB Lange 耦合器可以用作正交合成器或者分配器。一个实践上的限制是那些耦合线会是相当地窄，姑且不论实现为图 7.56 所示的用键合线来实现连接线的困难(即使是使用与附加金属层相联系的通孔进行连接也是如此)。一个能补偿这些问题的优势是，Lange 耦合器可以工作在比支线混合型耦合器能提供的带宽要高出 10%或者 15%的带宽。像普通的耦合线耦合器那样，一个 Lange 耦合器的典型有用的带宽超过二倍频程[56]。该属性说明了尽管有那些制造上的挑战，Lange 耦合器依然是那么受欢迎。

下面这个 12GHz 耦合器(采用仿真器 Sonnet Lite 9.51)的仿真以更详细的方式突出了传统的 Lange 耦合器的性能。正如从图 7.58 中能看到的，在一个频率由低于 8GHz 到高于 16GHz 的范围内，耦合是接近于 3dB 的。在同样的频率范围内，输入返回损耗超过 18dB。

鉴于给予紧耦合的所有这个关注，很容易形成这样一个错误的印象，即实现最大的耦合总是一个最重要的目标。因此注意到下面这点是十分重要的：在许多场合下，例如 SWR 的测量，或者对放大器输出的取样以测量功率或者去完成一个反馈回路，不总是需要最大耦合。因此，应该可以自由地来考虑采用线长短于 $\lambda/4$ 的一个传统反向波耦合器，特别是当空间是一个重要的工程考虑的情形下。

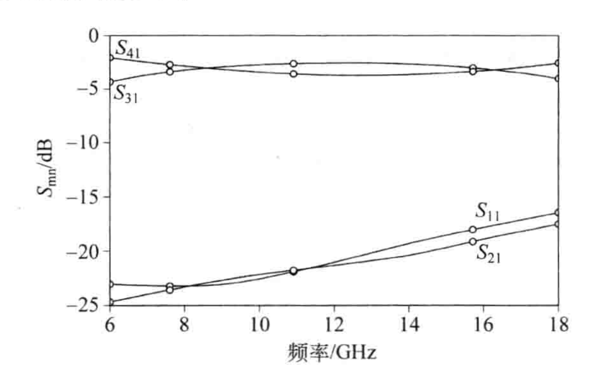


图 7.58　一个 12GHz Lange 耦合器的模拟得到的性能

7.6.5　宽带和尺寸缩小技术

我们已经研究了耦合器中广泛地使用传输线段，其长度用波长的分数来表示，这一般是意味着窄带工作。这种性质有时是一个长处而不是一个限制，这是由于它能提供对频带外的噪声和干扰的非有意识的滤波作用，这是一个依靠频率的耦合的副产品。然而，也有许多场合(特别是在仪器的应用上)，宽带工作的是十分希望的。因此，考虑能扩展耦合器的工作频率范围的方法，是有价值的。

最为通用的宽带技术，即使用多段技术，在概念上是简单的，但是实现起来却有点困难。就像步进阻抗变换器，通过级联多段能够使更宽带的工作成为可能，级联多段中每一段在总

[56] 这个说法适用于耦合的大小，而不是正交相位的保持。

的性能中承担一小部分负担。见图 7.59，这个方法也可以应用在耦合线上，提供在数十倍带宽内的好的耦合系数和定向性。

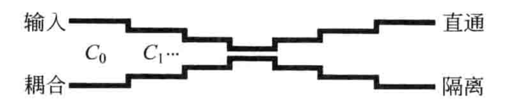


图 7.59　多节耦合器(所示的是对称版本，相对大小不成比例)

就像我们处理传统的单段反向波耦合器，考虑一下一个波从输入端口流向输出端口。这个波在一条耦合线中引起了一个反向传播的波。在这个特定的例子中，当这个波经过各个相互耦合的线段时诱发了一连串的反向传播的波。例如，一个波形在第零个线段中传播时，在耦合线中产生了一个耦合系数为 C_0 的反向传播波，然后当正向波接着进入下一线段时，产生耦合系数为 C_1 的另一个反向波，以此类推。当所有被诱生的波反向流向耦合端口时，它们叠加在一起构成了总的耦合信号。

现在我们已经看到反向波耦合器实际上依赖于反射来工作。弱的耦合产生小的在模阻抗上的差别，因此它产生小的反射。如果进一步假设级联这种弱的耦合器并不明显地改变这种弱反射一般性质，可以再次引用这个小反射近似，因而来估计总的耦合为延时耦合的简单的加权和。

作为开始，产生一个在弱耦合极限下的耦合因子的近似。从公式(103)开始，可以得到以下这个表达式：

$$\begin{aligned}\frac{V_3}{V_1} &= \frac{(\mathrm{j}\tan\theta)C_\mathrm{F}}{\sqrt{1-C_\mathrm{F}^2}+\mathrm{j}\tan\theta}\\ &\approx \frac{(\mathrm{j}\tan\theta)C_\mathrm{F}}{1+\mathrm{j}\tan\theta} = \frac{(\mathrm{j}\sin\theta)C_\mathrm{F}}{\cos\theta+\mathrm{j}\sin\theta} = [(\mathrm{j}\sin\theta)C_\mathrm{F}]\mathrm{e}^{-\mathrm{j}\theta}\end{aligned} \tag{121}$$

这里，跟以前一样，C_F 是最大的耦合因子。注意所做的仅有的一个近似来自弱耦合假设(C_F 相比 1 是小的)。除了那个近似外，方程(121)依然适用于任意大的带宽。(再一次，这是在原则上是如此。)

下一步是，注意到第 N 段对耦合端口的贡献是被延时了 $2N\theta$，这是因为当前向波传输到第 N 段时已经有一个 $N\theta$ 的延时，而另一个 $N\theta$ 的延时是由于诱生的反向传输回来的波传输到耦合端口。将所有这些反向传回来的波叠加在一起，得到

$$\frac{V_3}{V_1} \approx \mathrm{j}\sin\theta[C_0\mathrm{e}^{-\mathrm{j}\theta} + C_1\mathrm{e}^{-\mathrm{j}(\theta+2\theta)} + \cdots + C_N\mathrm{e}^{-\mathrm{j}(\theta+2N\theta)}] \tag{122}$$

其中耦合段的总数为 $N+1$。

注意到公式(122)的方括号中的被加项是具有傅里叶级数的形式，我们可以看到(就像在其他的例子中，在那些例子中我们采用了小反射近似)系数 C_n 可以被选择成来产生一系列的电路特性。例如，在这里耦合被选择成可以使响应最大平坦化或称为等涟波[57]。

[57] 在这种结构与多节阻抗变压器之间存在着一个自然的特别紧密的联系。对于这个议题的一个很好的总结，参见 Inder Bahl 和 Prakash Bhartia, *Microwave Solid State Circuit Design*(微波固态电路设计)，Wiley, New York, 1988. 也可参考 Matthaei, Young, and Jones, *Microwave Filters, Impedance Matching Networks, and Coupling Structures*(微波滤波器、阻抗匹配网络与耦合器结构)，reprinted by Artech House. Dedham, MA, 1980. 称为 *MYJ* 或"(大)黑书"，这本经典著作综合性地描述了直到 20 世纪 60 年代中期有关这本书标题所涉及的领域的最近期的发展状况。

如图 7.59 所示，在不同段之间的过渡通常是斜接而不是突变相接的。就像在单段的耦合器中，通常每段的长度为标称的 $\lambda/4$（在频带的中心频率点上测量）。就如在单节耦合器那样，奇偶模相位速率的不等使用微带线制作的多节耦合器的性能变差。如果要用微带线实现合适的宽带性能，对这些效应进行的补偿（例如像以前介绍过的那样，用集总的电容器）是必需的。

Wilkinson 分配器通过级联可以做成在相似的宽的频率范围内工作，其频宽是十倍或更大，而不是由单独的一级所能提供的二倍频范围[58]。它可以被看成是 Wilkinson 分配器和一些阻抗变换器的结合。就像与多段阻抗变换器，小的反射近似允许我们来理解，采用越多的段就可以实现越大的带宽。

在如图 7.60 的示意电路中，每一段的长度为 $\lambda/4$。在实际的版图中，这些臂几乎绝不会是如示意图所示的直的与互相平行的。代之以是一般采用半圆弧来实现。之所以有这种选择是出于至少两个考虑。一是这种几何形状改进了宽长比，特别是在多段情况下。其二是平行线同时也是耦合线，而这种耦合可能会干扰分配器的正常工作。

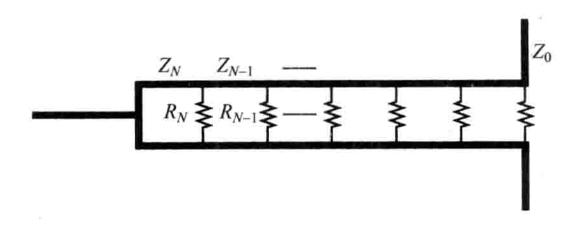


图 7.60　多节 Wilkinson（威尔金森）分配器（形式示意图）

对于分支线耦合器，基本的想法是相互耦合的几个谐振系统一起产生了在原来的，没有耦合时的谐振频率之上与之下的谐振。初始谐振频率和耦合强度的适当选择可以扩大工作频宽的范围。事实上，正是这种机制使得 Lange 耦合器具有令人惊异的高带宽。尽管这种现象在放大器与滤波器实际的场合下对工程师是非常熟悉的，同样的内在考虑可以应用到耦合的问题上。例如，可以将分支耦合器级联起来（图 7.61）来相当可观地扩大它们工作的频率范围。遗憾的是，对于这种分支线耦合器的通用设计，并不存在着简单的公式，仅仅有些算法[59]。

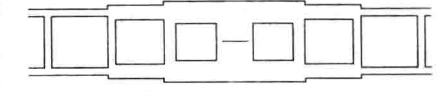


图 7.61　多节分支线耦合器

电路图在那些臂长还是对应于中心工作频率的 $\lambda/4$、而传输线阻抗的分布和线段数目的选择决定了带宽和通带内的纹波。就像与单个分支线耦合器和许多我们已经给出的其他例子，我们可以用一个 π 形的 *CLC* 网络来代替每一个 $\lambda/4$ 臂，以此来实现宽带多节正交混合变换器的集总元件版本。

---

[58] Seymour B. Cohn, "A Class of Broadband 3-Port TEM Hybrids"（一类宽带三端口 TEM 混合耦合器），*IEEE Trans. Microwave Theory and Tech.*, v. 16, February 1968, pp. 110-118.

[59] O. Maraguchi *et al.*, "Optimum Design of 3dB Branchline Coupler Using Microstrip Lines"（采用微带线实现的 3dB 分支线耦合器的优化设计），*IEEE Trans. Microwave Theory and Tech.*, v. 31, August 1983, pp. 674-678.

采用级联的一个折中后果是增加了尺寸。为了降低这个后果的成本，可以采用如图 7.62 所示的结构[60]。这个想法是采用锯齿状(或弯弯曲曲状)的几何形状线来压缩耦合段的长度。有效的电长度处于为线所占据的水平尺度与沿着锯齿边的总长度之间。就像人们可以预期到的，对于这种耦合器的精确设计，没有解析公式存在。然而，如果一个人从三段步进的阻抗变换器的原型出发(这个原型设计本身就不那么容易，但还是有可能的)，那么一个初始的锯齿型设计是可以通过一个简单的基于经验的步骤实现的[61]。对于那个间距最小的线段，这样来选择线段的长度以使得总共的锯齿状长度和原型的一样长。对于第三段(即有最大间距的那段)，假设边到边的距离是在锯齿边的中央来量度的。然后，只需选取中间段的线间距和长度的缩短为第一和第三段的值的几何平均值。再根据需要，通过仿真和设计改进来汇聚到一个可接受的设计。在一个 1～8GHz 的频率范围内，一个设计原型给出了约 10dB 的定向性，具有 1.5dB 纹波的 19dB 耦合度，以及最坏情况的回波损耗 16dB[62]。尽管这种性能水平本身还没有达到商业产品的标准，该设计代表了一个可信赖的第一轮设计过程，依此一个适当的耦合器可以经过数次迭代重复而得到。



图 7.62　Schiffman“鳄鱼齿”缩小了尺寸的宽带耦合器

最后，重要的是注意到，这些解决方案(交叉多指，曲折线等)中的许多之所以需要在很大程度上是因为要坚持用简单的平面方法实现。如果有额外的金属层可以利用，那么宽面耦合结构(而不是侧边面耦合)会以容易得多的方式提供紧耦合。确实，比如说通过使用(采用)宽面耦合的螺旋电感，尺寸可以大大缩小。

7.7　总结

本章介绍了微带线传输线的基本特性，以及用于估算电路常数和寄生参数的大量近似方程和经验规则。这些近似的许多有助于表征和缓解电抗的不连续性，这些不连续性发生在弯折处和其他线段间或者是在连接器和传输线之间的过渡。我们考虑了将适当设计的传输线线段用作电容、电感和谐振器。并且给出了对耦合器、分配器和合成器的详细的讨论(也对耦合器、分配器和合成器作了详细讨论)。

7.8　附录 A　一些有用的电感公式

我们在前面的一章中已经零零碎碎地介绍了一些电感的公式。我们现在要提供另外一些适用于经常遇到的几何形状的公式。下文中，除非特别说明，这些表达式仅严格适用于直流情况。在高频情况下，由于趋肤深度的减小而引起的导线内部磁通量贡献减小，电感值会

---

[60] 一如 S. Uysal 在下面这本书中所报告的：*Nonuniform Line Microstrip Directional Couplers and Filters*（非均匀线的微带线定向耦合器与滤波器），Artech House，Dedham，MA，1993.

[61] 见 Dana Brady，“The Design，Fabrication and Measurement of Microstrip Filter and Coupler Circuits”（微带线滤波器与耦合器电路的设计、制造与测量），*High Frequency Electronics*，July 2002，pp. 22-30.

[62] 同前 Brady。

下降一些。幸运的是，内部磁通量在总的直流电感中仅占很小的一部分（比如说，小于5%），因此，高频电感值的减小并未对整体的电感值产生很大的影响。然而，有必要通过弄清楚这些推导中都使用了哪些假设，从而避免令人不悦的惊讶。

在以下几节中，传输线效应是被忽略了的，因此这些公式仅适用于导体的尺寸与波长相比是较小的情况。

7.8.1 平板金属和圆导线

在第6章，我们给出了一个电流薄层（即只考虑面电流，译者注）的电感公式。这里再重述一遍这个公式，这样，所有的电感公式都放在一个地方以便于读者易于查找。

$$L_{\text{sheet}} \approx \frac{\mu_0 l}{2\pi}\left[0.5\ln\left(\frac{2l}{w}\right)+\frac{w}{3l}\right]=(2\times10^{-7})l\left[0.5\ln\left(\frac{2l}{w}\right)+\frac{w}{3l}\right] \quad (123)$$

详见6.5.2节。

圆形导线的直流电感如下[63]

$$L \approx \frac{\mu_0 l}{2\pi}\left[\ln\left(\frac{2l}{r}\right)-0.75\right]=(2\times10^{-7})l\left[\ln\left(\frac{2l}{r}\right)-0.75\right] \quad (124)$$

7.8.2 单圆环

对于单圆环的一个有用的近似表达式（很有用的一个近似公式）如下：

$$L \approx \mu_0 \pi r \quad (125)$$

这个公式告诉我们，半径为1mm的圆环的电感值约为4nH。

下面这个方程考虑了导线的直径不为零，与式(125)相比更为准确[64]：

$$L \approx \mu_0 r[\ln(8r/a)-2] \quad (126)$$

式中，a 为导线的半径。从式(126)中，我们可以知道，式(125)仅在 r/a 为20的情况下严格成立。

再进一步做近似，我们可以把式(125)扩展到非圆形的情况，认为面积相同的所有环路的电感值都大致相同，而不必考虑其形状。因此，方程可写成

$$L \approx \mu_0 \sqrt{\pi A} \quad (127)$$

式中，A 为环路的面积。详见6.5.2节。

7.8.3 平面螺旋线

为了保持本文研究平面结构的初衷，下面要介绍一个普遍用在PCB板（或者集成电路IC）中实现小值电感的常用的几何形状。圆形、八边形、六边形和方形的平面螺旋电感均有用到。电感值和可达到的 Q 值与形状呈二阶函数关系（尽管有很多相反的说法），因此工程师们可以毫无顾忌地采用他们喜欢的形状。方形螺旋线的版图结构最简单，因此最受懒惰的工程师们欢迎（笔者也不加掩饰地属于这类人）。

---

[63] The *ARRL Handbook*（*ARRL* 手册），American Radio Relay League，1992，pp. 2-18. 只要导电平面之间的距离至少为其长度1～2倍，它们之间的邻近效应可以忽略。

[64] Ramo，Whinnery，and Van Duzer，*Fields and Waves in Modern Radio*（现代无线电中的场与波），Wiley，New York，1965，p. 311.

表 7.5　电感公式的系数

形状	c_1	c_2	c_3	c_4
正方形	1.27	2.07	0.18	0.13
六角形	1.09	2.23	0.00	0.17
八角形	1.07	2.29	0.00	0.19
圆形	1.00	2.46	0.00	0.20

所有这些形状的公式可以写作一个统一的形式

$$L=\frac{\mu n^2 d_{avg} c_1}{2}\left[\ln\left(\frac{c_2}{\rho}\right)+c_3\rho+c_4\rho^2\right] \tag{128}$$

式中，n 为圈数，d_{avg}为内径和外径的平均值，ρ 为填充率，定义为

$$\rho \stackrel{\text{def}}{=} \frac{d_{out}-d_{in}}{d_{out}+d_{in}} \tag{129}$$

从上式中，可以知道为什么用填充率这个词是恰当的：当电感线圈充满了整个空间时，ρ 趋近于 1；当电感的中间空间变得越大时，ρ 趋向于 0。

各种 C_n系数是几何形状的函数，四种最具代表性的形状的系数列在表 7.5 中[65]。作为一个极佳的近似，系数 c_1 为版图中电感所占的面积比上能将这个面积内接的最大圆的面积。系数 c_2 为主要项，c_3 和 c_4可分别看成一阶和二阶修正因子。当四项因子均用到时，这些方程一般可精确到百分之几(误差几乎绝不会超过 5%)，因此不需要进行全波电磁场仿真来估算这些结构的电感值。

在需要采用其他正多边形的少数场合，可以使用下面这个解析公式：

$$L\approx\frac{\mu n^2 d_{avg} A_{out}}{\pi d_{out}^2}\left[\ln\left(\frac{2.46-1.56/N}{\rho}\right)+\left(0.20-\frac{1.12}{N^2}\right)\rho^2\right] \tag{130}$$

式中 A_{out}是外部总面积，N 是多边形的边数。该公式仅是式(128)的另一种表述，即用解析近似式来计算系数 c_1，c_2 和 c_4。系数 c_3 则设为 0，当规则多边形的边数大于 4 时，该假设是一个不错的近似。该解析表达式的准确性仅比之前给出的列表的式子低 1%～2%。

通过用趋肤效应公式来近似地计算平面螺旋电感的有效电阻值，可以粗略地估算其 Q 值。这个公式并不是十分恰当，因为它忽略了一个给定的线圈产生的场对相邻线圈中的电流分布的影响。因此这种估算方法最多是相当粗略的。

一般来说，中空的电感 Q 值最高，因为最内圈的那些线圈并不贡献多少磁通量却引起了相当大的电阻。因此，把内圈的线圈去掉总的说来是个好主意。尽管并没有简单的规则来指出什么样的结构在所有情况下都是最佳的，一个恰当的原则是外径和内径的比值为 3∶1。幸运的是，最优的条件是相对平坦的，因此这个原则在大部分实际情况下均是满意的。

除了串联电阻之外，人们也关心自谐振频率。线圈本身的寄生电容会引起谐振，这个电容主要因为连接线圈中心的线与电感线圈其他部分相互重叠而造成的。该重叠电容可以简单地从平板电容的公式估算得到。线圈之间的电容通常可忽略，这是因为这些单项电容均在互相串联后出现在电感的两端(因而总电容很小，译者注)。

---

[65] S. S. Mohan *et al.*, "Simnple Accurate Inductance Formulas "(一些简单而准确的电感公式)，*IEEE J. Solid-State Circuit*, February 2000.

7.9 附录B 边缘修正公式的推导

威尔·鲁滨逊(Will Robinson),危险啊——前面有积分,奶酪和清风!

边缘电容的严格计算是相当困难的。当需要得到精确答案或者几何结构很复杂时,最佳的实际选择通常是采用数值方法。遗憾的是,这个途径通常模糊了对设计的内在理解。作为对那些有用的数值方法的补充,这里给出了一个解析表达式,由于它的简单性和普适性,其不准确性也许能被接受。虽然它的推导或许不能在鸡尾酒餐巾上完成,但最终结果确实是这样的简单、紧凑(这在其在方程(12)中结合进去就变得清楚了)。

从事场研究的理论工作者已经设计出许多巧妙的策略来完成我们所要寻找的方法。我们将要采用的方法则是从费曼(Feynman)的 *Lectures on Physics*(物理学讲义)中的许多精彩的章节中的一章得到了启发,特别是“最小作用量原理”(The Principle of Least Action)[66]。在那本书中,费曼指出那个十分有用的最小值原理可以为一些古老的问题提供新颖而简练的解答。比如说,如果你忘了两个并联电阻的电流分配定律,那么你可以从电流必然要这样来分配以达到最小总功耗这一原则中得到。任何其他的电流分配都会使总功耗增加(你可以自己试试看)。同样是满足这个极小原理的为电压分配定律。

类似的,如果我们的任务是要推断出两个导体之间的电场,知道电荷的分布是要取得系统总能量的最小化这件事是很有价值的,同时也要记住与一个电荷分布,只存在着一个唯一的电势分布。因为(对于一个给定的电压)能量与电容成正比,从这个最小值原理可以推断出正确的电势分布是所有可能分布中的一个,它能最小化计算得到的电容[67]。我们这样来利用这个观察:先提出一个合理的电势分布函数的形式,计算它所隐含的电容值,然后选择参数(如果有的话)来使这个电容值最小化。费曼的最小值原理然后就说我们已经产生了一个基于这个特定的猜解(即使它是错的)的最佳的对实际分布的近似。而且,我们知道近似值的误差总是正的(因为我们的估算公式总是过高估计了真实的电容值)。

作为一个开始,令两个计算储存在一个电容器中的能量的公式相等:

$$\frac{1}{2}CV_0^2 = \frac{1}{2}\varepsilon\int_{\text{Vol}} |\nabla V|^2 \,\text{dVol} \tag{131}$$

式中,∇V 是电压 V 的梯度(记得电场等于电压梯度的负值)。上式中左边的项是来自普通的电路理论,右边的项则是来自场的理论。

接下来(这是棘手的部分),猜一个对于电势的合理的形式。为了便于猜解,首先看一下我们的结构,如图7.63所示。电场线被理想化为完全是垂直的,直到到达上部板的边缘,然后电场线逐渐向外弯曲,而且是愈来愈厉害,直到电场线在边缘之外 H 处成为完全的半圆圈。沿着图中所示的径向线——其与接地面呈 θ 角——假设随着半径 r 从0增加到 H 时,电势以某种形式增加。再进一步假设(相当有想象力),对于给定的 r,随着角度 θ 从0增加到 $\pi/2$,电势由0开始呈线性增加。同时假设当 $r>H$ 时,电场储存的能量可忽略。后面这个假设避免了一个尴尬的预测:当半径趋于无穷大时,电势会大于外加的电压 V_0。这也使

---

[66] Volume II,Chapter 19(Addison-Wesley,Reading,MA,1964)

[67] 同样这个最小值原理也可以用来得到电感的公式。

我们低估了储存的能量。这个误差至少是在正确的方向上的，以来抵消系统性的过估计，而这个过估计是采用错误电势分布（尽管有可能过补偿）的方法所固有的。

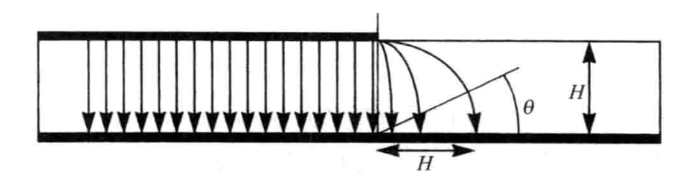


图 7.63　非常近似的为了估算边缘电容的场的分布（侧视图）

最后，假设两个平板为无穷薄。

给定了这些假设（就是这些了），可以假定一个以下形式的近似电势函数

$$\widetilde{V}(r,\theta)=V_0\left(\frac{r}{H}\right)^k\left(\frac{2\theta}{\pi}\right) \tag{132}$$

式中参数 k 的值将在后面确定。V 上波浪符表示它是一个假设的以及近似的电势。读者可以验证该方程满足之前描述的条件（但是并不是对所有相关的边界条件，不然的话，这个猜解就是正确的解了）。注意，我们忽略了 z 轴方向的变化（垂直于纸平面，在平面之外）。

得到了电势函数之后，剩下的工作就是数学运算了：

$$\widetilde{C}=\frac{4\varepsilon}{\pi^2H^{2k}}\int_{\text{Vol}}|\nabla\widetilde{V}|^2\mathrm{dVol}=\frac{4\varepsilon W}{\pi^2H^{2k}}\int_{\text{Area}}|\nabla\widetilde{V}|^2r\,\mathrm{d}r\,\mathrm{d}\theta \tag{133}$$

$$\frac{\nabla\widetilde{V}}{\widetilde{V}}=\frac{k}{r}\boldsymbol{r}+\frac{1}{r\theta}\boldsymbol{\theta}\left(\text{或}\nabla\widetilde{V}=\frac{\nabla\widetilde{V}}{r}\left(k\boldsymbol{r}+\frac{1}{\theta}\boldsymbol{\theta}\right)\right) \tag{134}$$

式中 $\boldsymbol{r}$ 和 $\boldsymbol{\theta}$ 分别是 r 和 θ 方向的单位向量。再将这些方程联立起来并且对 r 从 0 到 H，θ 从 0 到 $\pi/2$ 进行双重积分之后，得到

$$\frac{\widetilde{C}}{W}=\frac{\pi\varepsilon k}{12}+\frac{\varepsilon}{\pi k} \tag{135}$$

式中 W 为线宽。

现在，要选择 k 的值来使估算的电容值最小化。将式(135)的一阶导数设为零，得到

$$k=\sqrt{12}/\pi \tag{136}$$

考虑到整个上面推算过程的近似本质，这个结果可以看成基本是 1（意味着我们完全可以从 $k=1$ 开始，而最后得到差不多的同样答案）。

将 k 的这个值代入式(135)，最终得到单位宽度的边缘电容的近似方程：

$$\frac{\widetilde{C}}{W}=\frac{\varepsilon}{\sqrt{3}} \tag{137}$$

注意到这个电容值与 H 无关。就得到上面结果的推导近似程度而言，这个近似是一个普适的边缘修正值，大约为 5fF/mm。就是说，任何一个开放式结构会大约贡献这个大小的每单位长度的边缘电容，至少达到鸡尾酒餐巾的精度[⑱]。上面的推导是假设介质为真空，但是

⑱　对于一个在无限大平面上的有限长的导体，这个说法是正确的。对于两个等尺寸的平板，那个普适的修正值正好是该值的一半，即大约 2.5fF/mm。但是每个边缘的长度延展仍为 $H/\sqrt{3}$（或者取整，大约为 $H/2$）。同时要记住我们忽略了 z 方向上的场变化，因此当 W/H 减小时，这个修正值变得越来越不准确。

我们仍然假设即使对于微带线，这个修正依然是有效的。

因为电容值正比于平板的有效面积 WL_{eff} 和平板距离 H 的比，因此，边缘电容等于其面积为 W 和 $H/\sqrt{3}$ 的乘积的理想平板电容。

但是且慢，你说方程(24)中包含的是 $H/2$，而不是 $H/\sqrt{3}$，我们这里说明是如何得到这个 $H/2$ 的：首先，我们知道我们提议的函数形式是错的(考虑一下 r 大时的行为)，因此，如果当 $r>H$ 时场真的可以被忽略，那么根据最小值原理我们知道我们的估计可能太高(或者说$\sqrt{3}$是太小了)。但是我们的估算值高了多少？我们并不知道(如果我们知道的话，我们就不会有误差了)。但是如果假设估算值并没有大错，可以任意地将分母值增加一点，取最接近的整数值 2。这样做有点太粗糙了吧，的确如此。我们能估计一下这样做带来的误差吗？表 7.6 给出了直径为 D，间距为 H 的圆盘形平板电容的修正因子。我们将修正因子定义为实际(或者说估算的)电容值与从物理方程中得到的电容值之比。第二列值由场数值求解得到，第三列假设我们用 $H/2$ 对半径进行了伸展(因此有效直径是 $D+H$)。

表 7.6　圆盘形平板电容

H/D	"准确"的校正因子	粗略的校正因子(用 $H/2$)	残余粗略计算误差(%)	粗略的校正因子(用 $H/\sqrt{3}$)	残余粗略计算误差(%)
0.005	1.023	1.010	1.2	1.012	1.1
0.01	1.042	1.020	2.1	1.023	1.6
0.025	1.094	1.051	4.0	1.059	3.2
0.05	1.167	1.102	5.5	1.119	4.1
0.10	1.286	1.210	5.9	1.244	3.3

与期望的一样，这些修正因子对小的平板间间距平板电容十分接近于 1，因此所有三个公式给出的答案相互间差别都不大。然而随着 H/D 比的增大，不考虑边缘效应的平板电容公式越来越大地低估了真实的电容值。当 H/D 比等于 0.1 时，真实的电容值比不考虑边缘效应的平板电容公式计算得到的值将近大 30%。用粗糙修正因子在同样平板间间距得到的残差低于 6%。这个紧紧跟踪的现象是令人鼓舞的，因为修正因子是从矩形结构中得到的，但是却同样适用于圆盘形结构。因此，我们有关一个普适的边缘修正因子的结论看起来还是合理的。

作为最后一点说明，如果用 $H/\sqrt{3}$来代替 $H/2$，在表 7.6 的特定情况下，变小一些，这可以从最后的两列数据中看出[69]。在一个归一化间距为 0.1 时，修正因子变为 1.24；它使误差减小到略微大于 3%，即等于将用粗糙修正带来的误差降低了相对于不考虑边缘效应时造成的误差(0.286，根据表 7.6 第二列，译者注)整整一个数量级。那么，使用哪一个呢？幸运的是，边缘所产生的影响为一阶效应的公式添加了一个二阶效应的修正项，因此这些修正项中的小误差在总体误差中占很小的分量。选择使用 $H/2$ 还是 $H/\sqrt{3}$(或者其他值)并不是一个很重要的问题，而且选择的依据可以由其他的标准来决定。我这个懒惰的作者大部分时间使用更简单的值 $H/2$，因为它可使得另一类能量最小化，即作者自己的能量。

---

[69] 对这组特定的数据组，$2H/3$ 甚至是个更优的值，在 $H/D=0.1$ 时几乎没有误差。

7.10 附录C　其他材料的介电常数

很重要的一点是不要把本书的焦点放在FR4这一事实来给读者一个其他材料不被用在微波电路中的(错误)印象。相反,FR4在“关键性的”微波电路中很少被使用。它的主要优点是低成本,使它在大批量的消费应用中越来越受欢迎。本书选择FR4为主要媒介是为了承认其流行性以及对大部分微波文献缺少关于FR4的信息所作的回应。然而,当我们在追求最佳性能时,其他介电材料要比FR4用得更多。表7.7提供了其他材料的介电常数。并不是所有的材料都适用于微波电路,但是它们可能是经常遇到的。

表7.7　其他电介质

材料	ε_r	材料	ε_r
Al_2O_3(96%)(铝矾土)	9.5	Mica	5.4
Al_2O_3(99.5%)	9.8	Mylar	3
AIN	8.7	Paper	2.7typ.
BaTiO3	～600	Plexiglas	3.45
BeO(99.5%);有毒!	6.6	Polyethylene	2.25
Diamond	5.5	Polystyrene	2.55
Fused silica	3.82	PTFE(Teflon)	2.1
GaAs	13	TiO_2	～100
Ge	16	RT/duroid	2.5
Glass(borosilicate)	4.8	Si	11.7

(来源:主要来自AppCAD(安捷伦技术公司))。

第8章 阻抗测量

8.1 引言

时域和频域的表征都能向我们提供关于一个系统的综合信息。频域的表征需要有能力在一个宽的频域范围内产生和测量正弦电压和电流。网络分析仪，不管是基于标量还是基于矢量的实现，都是这类仪器的一个例子。

一个替代的方法是利用一些时域方法来表征一个系统。这一类的主要工具是时域反射计（Time-Domain Reflectometer，TDR），其本质是一个小型的雷达系统。TDR 朝向被测试的器件发出一个脉冲信号（主信号），然后再观察对这个信号的回应。相对于主信号的反射时间不同表明了不连续处的位置，而反射回来的脉冲的波形又包含了不连续处的一些重要信息。所以，利用反射计，就可以很迅速地得到电阻性的和电抗性的不连续的位置和特性（以此来估算如何对它们进行校正）。网络分析仪也可以提供这方面的信息，不过需要更多的工作量。

本章中，网络分析仪和 TDR 我们都会考虑到，先从后者讲起。

8.2 时域反射计

时域反射计有两个主要的应用，找到阻抗不连续的位置，并得到该不连续的特征。这些特点使得 TDR 可以直接应用于故障的矫正，还可以评估任意补偿的质量。

8.2.1 定位不连续

如图 8.1 所示，一个 TDR 由两个模块组成：一个脉冲产生器和一个示波器。脉冲产生器产生一个快速上升的阶跃信号，并将其作为被测试器件（device under test，DUT）的输入。同时这个信号的一部分被用来加到示波器，使得示波器的扫描与阶跃信号同步。那个脉冲产生器的同步信号（见图 8.1，译者注）则被定时为可以保证示波器显示出阶跃信号上升沿之前一段时间和上升沿之后一段时间中的电压变化情况。

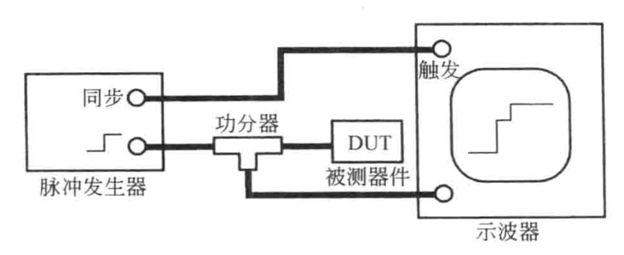


图 8.1 时域反射计（TDR）

脉冲信号的上升时间决定了信号的频谱，也因此决定了 TDR 能进行有效表征的带宽。类似地，示波器的带宽必须与想要表征的带宽一致。一个经验的准则是阶跃信号的－3dB带宽与其10%～90%上升时间成反比：

$$f_{-3\mathrm{dB}}t_r \approx 0.35 \tag{1}$$

这个关系，虽然严格说来是只对单极点系统成立，提供了估算对 TDR 系统性能要求的一个方法。例如，假设我们希望表征一条传输线，工作频率直到 10GHz 的特性。利用我们的经验准则，可以算得 TDR 的上升时间必须短于 35ps(皮秒)。最快的商用 TDR 可以表征系统超过 50GHz，这意味着上升时间要小于 7ps[①]。

入射脉冲的上升时间不仅决定了系统可以被表征的带宽，同时也决定了表征的空间分辨率。假若一个脉冲信号在距离信号源 x_d 处被一个不连续反射回来，那么信号往返于信号源的总时间是：

$$t_{\mathrm{prop}} = \frac{2x_\mathrm{d}}{v_{\mathrm{prop}}} \tag{2}$$

可以得到：

$$x_\mathrm{d} = \frac{t_{\mathrm{prop}}v_{\mathrm{prop}}}{2} \tag{3}$$

其中 v_{prop}是传播速度。显然，如果脉冲的上升时间太慢，反射信号会被上升沿所掩盖。粗略来说，空间分辨率约等于上升时间内的传播距离。最早的反射计是被用来定位很长的电缆线的故障位置，能够判断 100m 或者更长距离范围内的断路或者短路的情况就足够了。给定光在典型的电缆线中的传输速度是在自由空间中速度的 60%～80%，那么相对应的延时约为 4ns/m。上升时间在几百纳秒量级，也就是带宽在几兆赫兹的脉冲，就可以满足上述的电缆线故障位置定位的应用了。当今最先进的设备上升时间可以达到亚 10ps，其定位不连续处的分辨率在自由空间为几毫米。这样的上升时间和相应的空间分辨率与现在典型的微波电路元件和模块的尺寸是相适应的。

要定位不连续的位置并不一定要知道传播速度，虽然从式(3)来看这是必须的。例如对于微带线，只需要用手指沿着微带线移动并观察 TDR 的波形。当因手指导致的波形突起与你要观察的不连续所导致的波形突起一致时，这个不连续的位置就在你手指的位置下(当然，当 TDR 的脉冲能量很高时，这个方法不适用)。

8.2.2 表征不连续性

TDR 如此有用的一个原因是因为它传递的信息不仅仅是不连续的位置。这一点可以通过反射系数和终断阻抗的关系直接得以理解：

$$\Gamma = \frac{Z_{\mathrm{Ln}} - 1}{Z_{\mathrm{Ln}} + 1} \tag{4}$$

其中 Z_{Ln}是归一化的负载阻抗，

$$Z_{\mathrm{Ln}} \overset{\mathrm{def}}{=} \frac{Z_\mathrm{L}}{Z_0} \tag{5}$$

① 这里我们没有包括在利用低温和超导技术的系统。具备亚皮秒(ps)上升时间的脉冲发生器在实验室的演示表明带宽将持续增长。

注意到反射系数一般来说是一个复数，拥有幅度和相位两方面的信息（或者说实部和虚部）。所以反射系数包含了因不连续引起的反射如何改变阶跃响应的频谱分量的信息。同时注意到Γ包含了负载阻抗的所有信息（如式(4)所示，可以通过Γ反推出Z_{Ln}）。同时，线性的假设允许我们将线性系统理论所有的有用工具都用来解决这个问题。特别是，得到阶跃的响应是系统理论中最古老与常见的应用，这也正好是TDR提供的信息。尽管我们会从正式的数学推导开始，我们还是可以通过很快地分析几个有代表性的例子，来提取出如何可以猜测出对这些及其他有实际意义的问题的正确答案的内在理解。

对于任一个输入的响应是输入激励和该激励引起的反射叠加的结果。反射信号就是Γ乘以入射信号。因此，反映体系总的输出和输入关系的传输函数可以写为

$$H(s)=1+\Gamma=\left(\frac{2Z_{\mathrm{Ln}}}{Z_{\mathrm{Ln}}+1}\right) \tag{6}$$

当我们在用这个表达式时，一定要注意该式的拉普拉斯逆变换只在时间大于信号传播的往返时间时才正确：

$$t_{\mathrm{prop}}=\frac{2x_{\mathrm{d}}}{v_{\mathrm{prop}}} \tag{7}$$

在这个时间之前，输入端的响应只是输入信号值本身（例如1V，如果我们假设激励是单位阶跃信号）。

利用这些关系，就可以直接得到几种常见情况下TDR的轨迹。例如在开路和短路负载的情况。在这两种情况下，归一化的负载阻抗分别为无穷和0，对应的$H(s)$的值是2和0。注意，这些值只在大于信号传播的往返时间时成立，因此对单位阶跃信号的响应如图8.2所示。如果是阻值介于上述两种极端情况中的电阻负载，阶跃响应的值也应介于0和2V之间。如果负载电阻小于特征阻抗，响应的终值应小于1V；如果负载电阻大于特征阻抗，则相应的终值在1～2V之间；如果等于特征阻抗Z_0，则将观察到没有不连续的情况。

现在考虑终端负载为电抗性时的阶跃响应。如果负载是个电容，则

$$\frac{2Z_{\mathrm{Ln}}}{Z_{\mathrm{Ln}}+1}=\frac{2}{1+1/Z_{\mathrm{Ln}}}=\frac{2}{1+sZ_0C} \tag{8}$$

这就是一个单极点低通滤波器的传输函数，其阶跃响应大家应该是熟悉的，见图8.3。对任意数量不连续的阶跃响应都可以通过这样的方法类似地确定。在图8.4中，我们忽略细节的推导（这可以作为留给读者的练习）而给出一些简单但实际的不连续的情况和与之相对应的TDR轨迹。

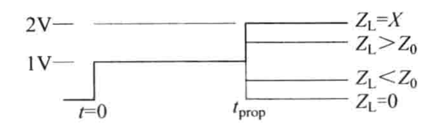


图8.2　对于开路、短路与电阻负载的理想化的TDR轨迹

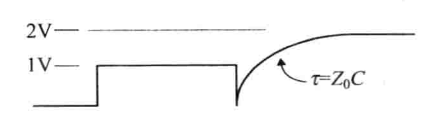


图8.3　为电容性负载终断的传输线的理想化了的TDR轨迹

TDR轨迹的形状可以通过纯物理的推断加上一点数学来得到。在所有的电抗的例子中，只涉及一个时间常数，因为我们只考虑了单电抗负载的情况。一个时间常数表明只有一

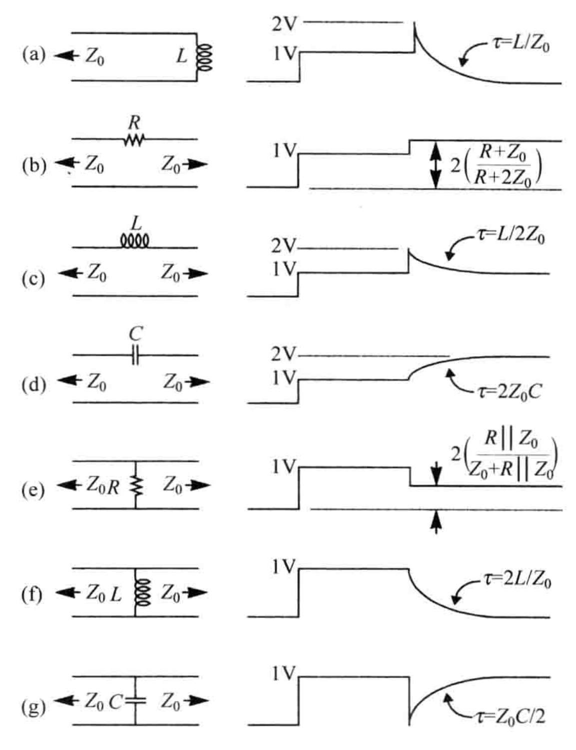


图 8.4　对于几种不连续性的理想化了的 TDR 轨迹(在所有情况下,输入幅度＝1V)

个指数因子。一个电感性的负载(图 8.4(a))在开始时呈现开路的性质,但最后处于短路状态。这两种状态之间指数转换的时间因子等于电感与它看见的有效电阻(此处为 Z_0)比值。在图 8.4(c)这个例子中,电感"看见"的总的电阻为 $2Z_0$(左右两边分别有一个 Z_0),且终值电压为 1V。

在图 8.4(b)中,包含了一个串联电阻的不连续,此时的阶跃响应必须突然跳起,因为此时的有效负载电阻是从不连续处往右看得到的值。此时为 $R+Z_0$。一个简单的电阻分压公式就可以得到图中所示的结果(记住,在开路情况下阶跃幅度是 2V)。

利用同样的分析方法可以得到图 8.4 中其他例子的 TDR 轨迹的结果。

实际上,观察到的 TDR 轨迹会与图中所示的理想曲线有所不同。主要的不同是由阶跃激励有限的上升时间引起的。我们可以把一个实际的阶跃信号看成是一个上升速度为无限快的理想阶跃信号通过低通滤波器的结果,那么实际得到的 TDR 轨迹可以从用一个阶跃响应的上升时间与实际阶跃信号上升时间相同的低通滤波器来推演出来。这个滤波器将减缓上升沿的速度,也是轨迹中锐利的转角变得圆滑。

8.2.3　参数提取

利用我们列举出的 TDR 的轨迹,一般可以从得到的阶跃响应中测量出小的电感和电容,甚至可以提取更加复杂的电路模型。要提取电路模型,我们需要清楚地考虑所有实际体系所具有的有限带宽会给波形带来什么样的影响。作为一个专门的例子,我们考虑图 8.5 中所示的因并联电容而引入的不连续。

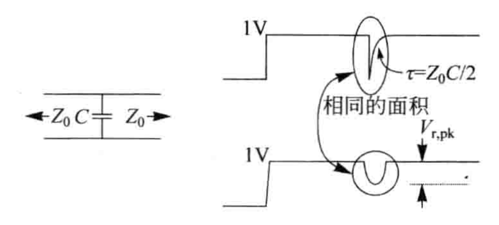


图 8.5　理想化了的与更为实际的因存在并联电容而形成的 TDR 轨迹

这个电路的反射系数为

$$\Gamma = \frac{Z_{\text{Ln}} - 1}{Z_{\text{Ln}} + 1} = \frac{1/(sCZ_0 + 1) - 1}{1/(sCZ_0 + 1) + 1} = \frac{-sCZ_0}{sCZ_0 + 2} \tag{9}$$

可以看到 Γ 在下面这个频率上有一个极点：

$$\omega = 2/CZ_0 \tag{10}$$

高于这个极点频率的频域分量将被由这个电容构成的低通滤波器有效地减弱。这个滤波作用就是图 8.5 中轨迹形状改变的原因——尖锐的转角变得模糊了，结果是图中下部轨迹的平滑的突起。尽管在形状上变得圆滑，但两个突起的面积相同（我们将在另一种测量方法中利用这个不变性）。如果这个电容不连续足够的小，那么这个滤波器的作用将可以被忽略。此时，反射系数可以近似地表示为

$$\Gamma \approx \frac{-sCZ_0}{2} \tag{11}$$

则入射信号和反射信号可以用一个导数近似地联系起来：

$$V_{\text{r}} = \Gamma V_{\text{i}} \approx \frac{-CZ_0}{2}\frac{\mathrm{d}V_{\text{i}}}{\mathrm{d}t} \tag{12}$$

V_{r} 的峰值与输入斜率的峰值成正比，因此

$$C \approx \frac{-2V_{\text{r}}}{Z_0}\left(\frac{\mathrm{d}V_{\text{i}}}{\mathrm{d}t}\right)^{-1} \approx \frac{-2V_{\text{r,pk}}}{Z_0}\left(\frac{V_{\text{i}}}{\tau}\right)^{-1} \tag{13}$$

其中，$V_{\text{i,pk}}$ 如图 8.5 所示，V_{i} 是输入阶跃信号的幅度，τ 是输入阶跃的时间常数。（我们已经利用了这样一个结论，即对于单极点系统阶跃响应斜率的最大值等于阶跃信号的幅度除以时间常数。）10%～90%上升时间约等于 2.2τ，所以也可以写成：

$$C \approx \frac{|V_{\text{r,peak}}|}{1.1Z_0}\left(\frac{t_{\text{rise}}}{V_{\text{i}}}\right) \tag{14}$$

利用相似的方法可以推导在串联电感引入不连续的时，电感的估计值为

$$L = 2Z_0V_{\text{r}}\left(\frac{\mathrm{d}V_{\text{i}}}{\mathrm{d}t}\right)^{-1} \approx \frac{Z_0V_{\text{r,pk}}}{1.1}\left(\frac{t_{\text{rise}}}{V_{\text{i}}}\right) \tag{15}$$

一个典型的这种不连续的波形如图 8.6 所示。这个方法可以提供令人惊讶的测量分辨率。假设一个 TDR 系统在上升时间为 15ps、幅度为 1V 的情况下，可以分辨的最小电压为 1mV，则其可以测量的最小电容和电感分别为 0.3fF 和 0.7pH！毋庸置疑的，要想利用其他方法对如此小的量进行测量是非常困难的。从这个计算中，我们可以清楚地发现，即使是相对而言不太敏感和较慢的 TDR 系统都有能力对电容和电感进行准确的测量。

另一种替代的测量方法

前面提到的方法是基于微分的。因为沿着一条线可能存在的各种各样的非理想情况都

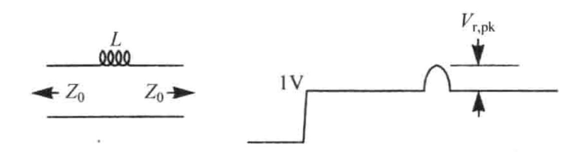


图 8.6　由于串联电感存在的比较接近实际的 TDR 轨迹

将在一定程度上影响到波形，所以这个方法很难如上面描述的不加修改直接应用到测量中。在这种情况下，利用基于积分的方法测量一般总是可以得到更为满意的结果。

依然从入射信号和反射信号的关系开始：

$$V_r = \Gamma V_i \approx \frac{-CZ_0}{2}\frac{dV_i}{dt} \tag{16}$$

我们可以说(注意，这里假设输入电压是单位阶跃信号，译者注)：

$$\int V_r dt = \frac{-CZ_0}{2} \Rightarrow C = -\frac{2}{Z_0}\int V_r dt \tag{17}$$

类似地：

$$\int V_r dt = \frac{L}{2Z_0} \Rightarrow L = 2Z_0\int V_r dt \tag{18}$$

对于大多数人来说，估计一个突起的面积比计算类似于斜率的参数要容易得多。所以，当突起是相当圆滑的情况下，计算面积的方法更为可取。

8.2.4　补偿

通过确定不连续的位置和类型，TDR 方法可以帮助你去设计补偿(如果需要的话)。因为 TDR 测量速度快，所以任何补偿设计的效果都可以很快地得到评估。作为一个专门的例子，考虑如图 8.7 中所示的斜接拐弯线。

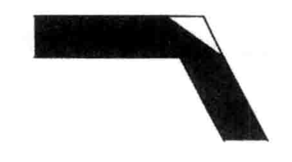


图 8.7　带有斜角的拐弯线

最佳的斜切量可以容易地通过 TDR 实验测量得到。只需要将拐角处一小块一小块地割掉，直到反射达到最小为止。如果是通过网络分析仪或者是槽线驻波比(SWR)测量法得到同样的结果，则会需要耗费更多的(或许是非常多的)工作。

一个不容忽视的问题是，某个不连续可能会掩盖其后不连续的存在或其大小。例如，一个大的串联电感(或者是一个大的并联电容)可能会减小 TDR 在该不连续之后的脉冲的带宽，因而使得表征出这个电感后面的不连续的机会降低。所以，正确的方法是首先修复离源头最近的不连续，然后再进行 TDR 测试，再修复好下一个不连续，如此往复，直到所有问题被解决为止。

8.2.5　TDR 总结

TDR 是对传统频域测量仪器的一个不可或缺的补充，它让我们可以在很短时间内测量出微波系统在较宽的频率范围内的特性。能够对不连续定位是 TDR 非常有价值的一个功能，另外它还能方便地在一个较宽频率范围内评估任意的补偿方法的性能。

1964 年 2 月的 Hewlett-Packard Journal(v. 15，no. 6)给出了一个非常好的 TDR 使用

说明。虽然40年过去了，但是基本的原理并没有改变。

8.3 开槽线

8.3.1 引言

自动矢量网络分析仪(VNA：vector network analyzer，或简单地网络分析仪：network analyzer)的发展给微波电路的表征带来了革命。通过计算网络在一个较宽频率范围所有的 S 参数值，网络分析仪给设计者全面地提供了电路行为的信息。这些信息若是通过手动的方法来得到是十分繁琐的。

在讨论VNA之前，我们先介绍一种被VNA在很大程度上已替代的工具——开槽线。回顾这个问题有几方面的原因：一是教育学上的，因为槽线向我们提供了直接观察波的反射和干涉现象等波动理论中精髓现象的机会。另一方面出于实用的考虑，因为开槽线利用了以相对便宜的仪器来测量高频下阻抗的一些现象。再一方面的原因是适用于VNA测量的一些重要校准事项可以由对开槽线的讨论自然地引入。最后，通过介绍利用开槽线进行精确测量需要耗费大量的人力这一事实可以说明为什么要发展网络分析仪的动机。

然后我们将详细介绍VNA，包括举例说明如何使用VNA。其中的一个重点是在确定和减少误差来源上，还要综合地描述校准技术，因为现代VNA的大部分功能来源于其表征并去除自身的一些误差。

这章的最后将阐述如何制作一个廉价的，能够在1～5GHz范围内测量阻抗的开槽线系统。

8.3.2 往昔：开槽线阻抗测量

在VNA发展之前，表征微波系统是一个繁琐的过程。首先考虑测量阻抗这样一个基本问题。在低频下，采用电桥测量是一个相对简单的方法，或者是给网络一个激励(如电压)然后测量对激励响应的电流。找到电压电流比是直截了当的，即使是我们需要知道两者之间的相对相位信息，以此来推导阻抗的实部和虚部：

$$Z = \frac{V}{I} = |Z| e^{j\phi} \tag{19}$$

但是随着频率的增高，情况不断地变得更加复杂。除了常见的仪器在高频下工作的困难外，还有夹具引入的严重问题——固定长度导体的阻抗会随着频率的增高越来越严重地干扰到测量。为了保证测量的是器件的真实性能，而非器件和连接线共同作用下的性能，测量时需要格外地小心。

优秀的工程实践上一个常出现的主题是将负担转换成财富(“这不是一个错误，而是一个特点”)。在上面提到的这个情况中，我们必须要承认这样一个事实：试图将夹具的阻抗减小到一个可以忽略的程度是徒劳的。所以，我们并不是尝试量化并除掉夹具对阻抗测量的影响，而是考虑负载阻抗对夹具的影响。为了理解为什么这种看法上的改变是有价值的，回顾一下，正弦电压信号的幅度只有对终端负载是匹配的传输线(或者波导)时，其幅度才不随位置的改变而改变。任何其他的不匹配都会引起反射，而这个反射波将与入射波周期性地形成或者是增强、或者是减弱的干涉，因而沿着传输线形成驻波。驻波的幅度和相位唯一

确定地依赖于负载阻抗和传输线特征阻抗(Z_0)之间的失配程度。通过对驻波的测量,再结合已知的 Z_0,可以计算得到负载阻抗 Z_L。这种阻抗测量方法的核心就是利用了阻抗和反射系数之间的双向唯一关系:

$$Z_{Ln} = \frac{1+\Gamma}{1-\Gamma} \tag{20}$$

其中 Z_{Ln}是归一化负载阻抗,

$$Z_{Ln} \stackrel{\text{def}}{=} Z_L / Z_0 \tag{21}$$

且反射系数是一个复数:

$$\Gamma = |\Gamma| e^{j\phi} \tag{22}$$

将沿传输线的电压写成入射分量和反射分量的和后,这个测量技术中的数学基础就变得清楚了:

$$V(z) = V_f + V_r = V_f(e^{-j\beta z} + \Gamma e^{j\beta z}) = V_f e^{-j\beta z}(1 + \Gamma e^{j2\beta z}) \tag{23}$$

这里,$z=0$ 是负载所在的位置,越接近信号源 z 值越负,β 是相位常数,$2\pi/\lambda$。

将沿传输线的电压幅度写成关于位置的函数:

$$\begin{aligned} |V(z)| &= |V_f e^{-j\beta z}||(1+\Gamma e^{j2\beta z})| \\ &= |V_f||(1+\Gamma e^{j2\beta z})| = |V_f||(1+|\Gamma| e^{j(\phi+2\beta z)})| \end{aligned} \tag{24}$$

其中 ϕ 是反射系数的相位角。从式(24)可以知道电压的幅度是周期性的。驻波的周期为 $\lambda/2$,所以最小值之间的距离所对应的相位幅度弧度值为 π,参见图 8.8。幅度的最小和最大值分别发生在指数项为-1 和$+1$ 的时候:

$$V_{min} = |V_f|(1-|\Gamma|) \tag{25}$$

$$V_{max} = |V_f|(1+|\Gamma|) \tag{26}$$

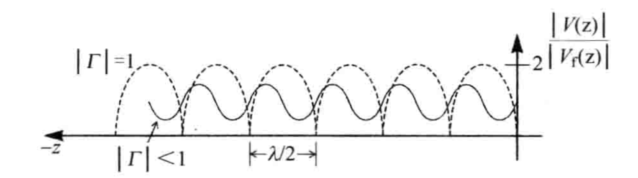


图 8.8　对于两个不同 Γ 值的幅度与空间位置关系的典型图

驻波比(SWR)的定义是最大幅度与最小幅度的比:

$$\text{SWR} = \frac{V_{max}}{V_{min}} = \frac{1+|\Gamma|}{1-|\Gamma|} \tag{27}$$

从式(27)可以清楚看出,通过测量 SWR 就可以计算出$|\Gamma|$。

为了完成测量,我们需要 ϕ,也就是 Γ 的相位。关键是要注意到幅度最小值发生在:

$$1+|\Gamma| e^{j(\phi+2\beta z)} = 1-|\Gamma| \tag{28}$$

或者等价地说:

$$\phi + 2\beta z = (2n+1)\pi \tag{29}$$

这里 n 是任意整数。所以,Γ 的相位可以通过如下公式计算得到:

$$\phi = (2n+1)\pi - 2\beta z \tag{30}$$

其中 z 是幅度最小值的位置(再次强调,在我们的坐标体系中,z 是一个负数)[②]。

在实际中,电学参考面 $z=0$ 的准确位置并不总是那么明显。所以,一般还需要另一个实验来得到这个参考面的相关信息。传统的(也是最简单的)方法就是将传输线尽可能良好地加以短路。显然,那些最小值就应该都为 0(这是理想的情况),同时这些零值点依然沿着传输线周期性地分布。任意一个 0 值点都可以看作是参考平面 $z=0$,尽管习惯上选择离短路处最近的那个点。选择一个这样的点,并记录下它的位置。同时记录下相邻两个 0 值点的距离(这个值应等于 $\lambda/2$),这样就可以以此计算相位常数 β。然后就是用需要测量的阻抗来替代这个短路。测量 SWR 来计算 $|\Gamma|$,同时记录下短路时和有负载阻抗时最小幅度位置的偏移。将离开负载的偏移记为负值。将这个值代入到式(30)来求解 ϕ。然后将式(22)代入式(20),来得到(归一化的)负载阻抗。真正的负载阻抗还需要将这个值再乘以 Z_0 来得到。

这种方法的绝妙之处在于并不要求夹具的长度小于波长,实际上,为了表征这些驻波,夹具的长度应该超过半波长(最好是能有几个半波长)。

这种测量方法需要知道电压沿导线的变化情况,这又要求我们能够通过物理路径接触这根传输线。因此,开槽线系统由一个用空气作为介质的传输线(或波导)组成,在其上切开一个可以放入探针的槽(这个探针就是一个简单的具有高阻抗的二极管检波器通过电容跟导线耦合起来)。这个开口和探针都需要仔细设计使其对电磁场的扰动最小。对于同轴线的情况,沿着长度方向在外层导体上的槽产生的影响最小,因为沿圆周方向并没有电流流动。这个槽的主要作用是使得单位长度的电容略有减小,从而导致特征阻抗略有增大。一个足够窄的槽将使这个影响保持在可以忽略的程度,并且保证辐射和辐射伴随着耗散达到最小[③]。同样地,可调整二极管检波器的耦合,使该检波器负载对测量的影响最小。

探针安装在一个带有刻度尺的游标上,这样沿传输线方向的位置是可以测量的(图 8.9 所示的是同轴线,开槽波导的情况与之相同)。在大多数开槽线中,探针进入传输线的深度是可以调节的,这样可以在检波的灵敏度与对电场分布的干扰之间做折中。幸运的是,探针的存在并不影响幅度等于零的位置(因为此处电场为零),因此,可将探针的位置调到高灵敏度处,以得到精确的数据。但是,探针会影响驻波的波形(幅度越大形变越大),这就会对峰值的幅度(与位置)的测量引入误差。幅度和位置的不对称程度,定性地反映了探针引入的扰动。

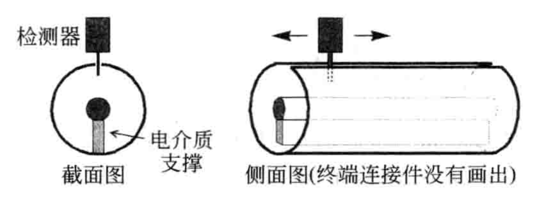


图 8.9　同轴开槽线

② 原则上讲,也可在测量中用最大值。但最小值更为显著些,因此对一个给定的幅度测量的不确定性,用最小值而不是最大值,可以得到较小的(或许是小得多的)定时(即相位)的不确定性。

③ 只要槽与导体壁的厚度可相比(或者窄于厚度),其行为就像一个截止频率足够高的波导。所以,辐射基本上总是可以忽略,以至很多教科书连辐射可能存在这一点都不提。

一个例子

在介绍完方法和设备之后，通过一个带真实数值的例子来阐明整个过程是很有帮助的。

第一步：确定参考面的位置。连接一个短路负载，同时找到最小值的位置（如果短路负载接得较好，且导线的耗散可以忽略的话，最小幅度应该接近于零）。可以自由地增加探针的深度，以得到更大的检测器的输出结果，从而提高确定零值点位置的精度。

假设在我们的例子中，最小值发生在 $z=-1\text{mm}, -121\text{mm}, -241\text{mm}$，同时注意到波长是最小值点之间距离的两倍，即 240mm。

第二步：将短路负载用需要测量的负载代替。将探针插入的深度减小以降低场分布的畸变程度（这可以通过波形的对称性来判断），同时要确认探针的输出是足够小而处于可测量的范围。记下 SWR 和新的最小值的位置。就像在大多数情况下，如果探针产生的输出电压与功率成正比，不要忘记通过对探针输出的最大值与最小值的比值开根号来计算 SWR。

在我们的例子中，假设测量的 SWR 值为 1.6，而且新的最小值位置在 $z=-41\text{mm}$，-161mm，-281mm。

第三步：从第一步中选择一个最小值的点作为初始点。计算在有负载情况下，这个点的坐标与相应的短路情况下坐标的偏移值。

这里，我们选择 $z=-1\text{mm}$ 为初始点（它最接近负载）。然后得到用于计算 ϕ 的偏移便为 $-41\text{mm}-(-1\text{mm})=-40\text{mm}$，给定波长为 240mm，我们计算得到 ϕ：

$$\phi=(2n+1)\pi-2\beta z=\pi-\frac{4\pi}{240\text{mm}}(-40\text{mm})=\frac{5\pi}{3} \tag{31}$$

这里任意地取 $n=0$

第四步：计算 Γ

首先从 SWR 的测量中得到 $|\Gamma|$

$$|\Gamma|=\frac{\text{SWR}-1}{\text{SWR}+1}=\frac{0.6}{2.6}\approx 0.23 \tag{32}$$

接着利用第三步中得到的幅度角来计算 Γ：

$$\Gamma=|\Gamma|e^{j\phi}\approx 0.23e^{j5\pi/3}=0.23\left[\cos\frac{5\pi}{3}+j\sin\frac{5\pi}{3}\right]\approx 0.115-j0.2 \tag{33}$$

第五步：用式(20)计算归一化阻抗：

$$Z_{Ln}\approx\frac{1+(0.115-j0.2)}{1-(0.115-j0.2)}=\frac{(1.115-j0.2)(0.885-j0.2)}{0.885^2+0.2^2}\approx 1.15-j0.486 \tag{34}$$

然后乘以 Z_0（这里我们假设为 50Ω）来最终得到要求的负载阻抗：

$$Z_L\approx 57.5-j24.3 \tag{35}$$

我们可以看见，这个频率下的负载阻抗等于一个电阻串联上一个不算大的电容。

这就（差不多）是我们要完成的全部工作。

从这个例子中，可以清楚看出利用开槽线的方法需要花费相当的气力才能得到在单一频率点下的阻抗特性，更不要说得到宽频域范围的情况了。这就是为什么今天这种方法不太常用的原因，尽管在毫米波电路中依然使用开槽线，因为在毫米波情况下，VNA 不是太贵了，就是根本就没有，或者是因为夹具的不连续会使得测量值不可靠。因为开槽线装置能很容易在二手市场上以低价购得，所以毫无疑问它们仍然是业余爱好者或者经费紧张的实

验室的最好选择。在 8.9 节我们将介绍一种更加廉价的选择——如何制造一个简单的基于微带线的“开槽”线装置。

误差源(及如何抑制)

机械上的瑕疵是误差的一个来源。例如，中心和外层的导体并不是完全圆柱形及没有完全同心，则阻抗就会与位置有关。同样的，如果探针夹持组件与中心导体不能保持相同的距离的话，那么就会引入一个与位置有关的误差。最后，为保证机械稳定的介质支撑也会不可避免地影响到电场分布。在图 8.9 中，这个支撑层在底部是连续的，但也常采用周期分布的柱状支撑来减小对场分布的扰动，只要能使这些支撑的间距不妨碍足够的机械支撑即可。不管怎么说，最好的开槽线是机械工程中最佳的例子，它们具有几乎完美的同心性。很多都装配有低回程间隙的游标，可以使得测量的精度高于 25μm。

另一种(而这通常是主要的)误差源是与探针的特征相关的。大多数的探针就是简单的二极管电路，具有近乎平方律关系的检波器。所以，它们的输出电压粗略地与功率成正比。参见图 8.10。图中的电阻 R_1 并不是开槽线探针本身的一个部件，它在示意图上出现主要是提醒我们探针取样的电压是特征阻抗约为 Z_0 传输线上的电压(假设失配是小的)。

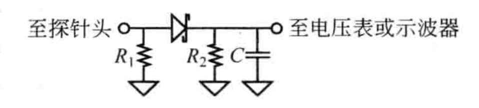


图 8.10　典型探针的原理图

为了粗略了解一下二极管作为功率探针的性能和局限，假设图 8.10 中的负载电容在 RF 频率下所产生的阻抗很低，以至于其上的电压降可以忽略。进一步假设二极管即使在 RF 频率下电流和电压仍然保持指数关系：

$$\begin{aligned} i_{\mathrm{D}} &= I_{\mathrm{S}}\left(\exp\left[\frac{qv_{\mathrm{D}}}{kT}\right]-1\right) \\ &= I_{\mathrm{S}}\left[\frac{qv_{\mathrm{D}}}{kT}+\frac{1}{2!}\left(\frac{qv_{\mathrm{D}}}{kT}\right)^2+\frac{1}{3!}\left(\frac{qv_{\mathrm{D}}}{kT}\right)^3+\frac{1}{4!}\left(\frac{qv_{\mathrm{D}}}{kT}\right)^4+\cdots\right] \end{aligned} \tag{36}$$

接下来，使二极管上的电压(在给定的假设下，这个电压与探针上电压是一致的)为正弦：

$$v_{\mathrm{D}} = V_{\mathrm{p}}\sin\omega t \tag{37}$$

因为非线性性，二极管的电流包含了输入频率的奇谐波分量和偶谐波分量。所有的谐波时间平均值都为零，所以仅有零频分量对二极管的直流电流有贡献。在式(36)的展开中，只有偶次方项才能产生直流分量，因此

$$\langle i_{\mathrm{D}}\rangle = \left\langle I_{\mathrm{S}}\left[\frac{1}{2!}\left(\frac{qv_{\mathrm{D}}}{kT}\right)^2+\frac{1}{4!}\left(\frac{qv_{\mathrm{D}}}{kT}\right)^4+\cdots\right]\right\rangle \tag{38}$$

从上式可以清楚地知道，那个二次项才提供与电压的平方成正比(也就是与功率成正比)的二极管的平均电流。所有其他项只对误差有贡献(这可以从遵守一个线性功率判断得到)，随电压的增加这个误差也越大。如果我们给一个任意的可以容忍的误差范围，比如说，四次方项的贡献可以有为二次方项贡献的 5%那么大，那么就会有以下关系式：

$$\frac{1}{4!}\left(\frac{qv_{\mathrm{D}}}{kT}\right)^4 < \frac{1}{20}\left[\frac{1}{2!}\left(\frac{qv_{\mathrm{D}}}{kT}\right)^2\right] \tag{39}$$

等价于：

$$\left(\frac{qv_{\mathrm{D}}}{kT}\right)^2 < 0.6 \tag{40}$$

因此，非常粗略地说，我们必须把二极管电压的峰值限制在如下范围：

$$v_D < \sqrt{0.6}(kT/q) \approx 20\text{mV} \tag{41}$$

虽然实际的二极管检波器在特性上变化很大，但是一般来说，当其输出电压超过5～10mV(这大致相当于输入功率在－20dBm左右)的时候，结果一般是不可信的。随着输出电压的增加，二极管迅速地偏离理想情况，这主要是因为此时高次项输出迅速增加。这个有用范围可以通过添加电阻负载来扩大，付出的代价则是降低了输出电平。要理解为什么添加负载电阻这样简单的技巧会有用，就要知道：在开路时，输出电压的峰值几乎可以等于输入电压在高幅度时的值，这是一个线性关系(因此很明显是亚二次)，在电路上添加一个负载时造成了一个介于短路(此时随电压增大电流急剧增加)和开路(此时电流不会迅速增加)之间的情况，所以可将有效范围显著扩大。实验中发现用470Ω左右大小两倍之内的电阻作为负载能取得最好的效果。添加适当的负载后，可以接受的输入功率被扩展了另一个10dB或者更多。然而，输出大小的降低导致了在灵敏度和精度之间的折中。

前面的讨论清楚地表明降低检波器两端的电压峰值可以提高精度。但是，在灵敏度一定的情况下，降低电压峰值意味着要使检波器测量的电压比也降低。如果我们测量导线上的最小电压和除最大值外的其他电压(并记录这些电压点的位置)，那么我们就可以准确地实现减小检波器动态范围的要求。这个其他电压可以在数学上跟最大电压值联系起来，因为我们精确地知道驻波的形状。特别的，可以证明有如下关系存在[④]：

$$\text{SWR} = \lambda/\pi\Delta \tag{42}$$

这里，对Δ的定义如图8.11所示。这种方法只需要探针能够精确测量的电压比为1.41∶1(与之对应的功率比为2∶1)。当要测量的SWR值很高时这种方法特别有用，此时最大电压与最小电压的比值很高。

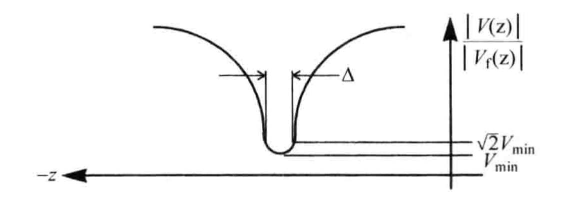


图8.11　对于高的SWR的另一种测量方法

一种明显的提高精度的方法是通过校准探针，解析地得到输入和输出之间的实际关系。但是工程师，他们总是很懒惰的(对不起，不要这么说，还是有效率的)，设计了一个聪明的步骤，完全不需要校正探针。因为SWR是没有量纲的，只需要测量一个比率就够了。考虑在信号发生器和开槽线之间插入一个经校准了的衰减器。将衰减调试到，比如说，最小值，然后将探针沿着传输线移动，找到最小值，并将那里的电压记录下来；然后移动探针来寻找最大输出值，再逐渐将衰减器的衰减系数调大，使得其与测得的输出最小值相等。则这个衰减系数正好等于V_{max}/V_{min}。注意，这个测量方法对探针没有任何要求，因为它用很容易实现的衰减器取代了校准非线性探针的需要。

---

④　参考，比如说：Terman and Pettit, *Electronic Measurements*, 2nd ed., McGraw-Hill, New York, 1952, p. 140. 这个方法也在 *Microwave Measurement* 中提到(MIT Rad. Lab. Ser., vol. 20), McGraw-Hill, New York, 1948.

最后，如果信号发生器可以产生一个被调制了的信号输出，那么灵敏度将有可能大幅度地提高。我们不是测量探针的直流输出，而是用一个在其后加上带通滤波器的解调器来得到输出。采用频率远高于所测系统的 $1/f$ 噪声拐角频率的调制信号，可以改进信噪比(SNR)，这样就允许提高检波器后级的增益，从而降低了对探针和传输线的耦合程度的要求。这个对扰动的减少，就提高了精度。

8.4 矢量网络分析仪(VNA)

8.4.1 背景

用开槽线得到每一个数据点都需要经过以下步骤：将频率设置到要求的值，用一个短路负载来定位新的电压参考平面(电压为零的点)，然后测量 SWR，以及在负载为被测器件(DUT)时定位电压最小值的点。矢量网络分析仪(VNA)自动地完成上面这个过程，还添加了很多其他的功能，使我们可以迅速而完整地得到一个微波电路在超宽频率范围内(例如从 50MHz～110GHz!)所有的 S 参数。

VNA 的核心是一个器件(例如定向耦合器)，它可以不可思议地将沿导线的信号分解为正向和反射分量。将信号分解为两个分量是十分有价值的，因为负载阻抗的测量就可以简化为对反射系数的测量，就如我们从开槽线的测量方法中已经看到的。类似的，对功率增益的测量只需要知道正向分量的比率，等等。所以 VNA 可以表征二端口系统的完整的 S 参数。

具有代表性的 VNA 方框图揭示了定向耦合器(或其等价物)起到的核心作用。如图 8.12 所示，一个频率合成器为网络分析仪提供输入。其输出功率和频率都是可控的。这个合成器输出的一部分被采样并作为入射信号，而剩下的部分被一对单刀双掷开关(SPDT)控制信号的去向。当这些开关处于图中所示的 F 位置时，DUT 被前向驱动，上方的定向耦合器提供一个次级输出，它相当于从 DUT 的端口 1 反射回的信号。同时下方的定向耦合器也提供了从 DUT 端口 2 输出的功率。

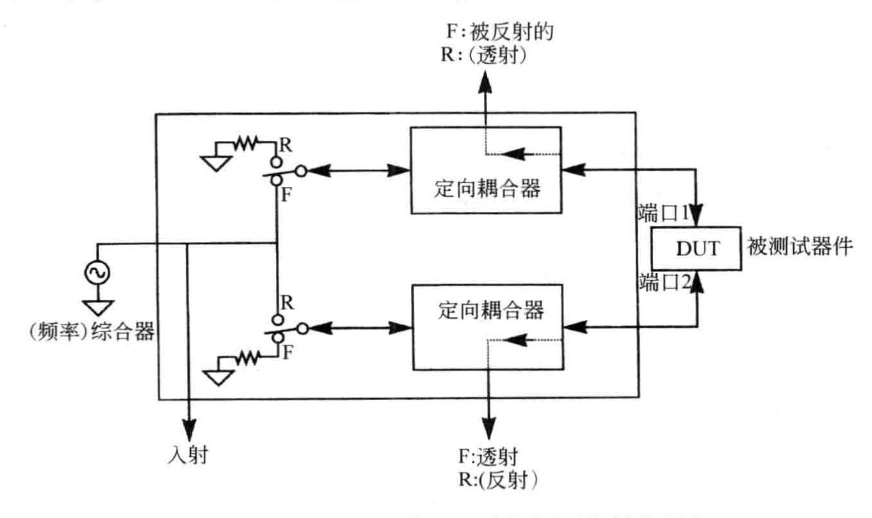


图 8.12 典型的 VNA 核心电路块方框图(简化版本)

为了测量反向特性，需要将这些开关放置在 R 处，使得 DUT 的端口 1 和端口 2 的作用倒过来。入射、反射和透射的信号都被送到接收/检波器(图中没有表示出来)。这个接收/检波器的作用是测量这些信号的幅度和相位，然后处理好这些数据后，在显示器上显示出来。

VNA 具有收发机的所有模块，以及其他的模块。注意到，VNA 不仅可以覆盖一个非常宽的频率范围(例如从 50 MHz 到 110 GHz，虽然在临近极限时性能有所下降)，VNA 同时还需要对幅度变化很大的信号进行准确的测量。要在这样大的范围工作，VNA 需要确定和修正尽可能多的内部和外部的误差源。现代的 VNA 采用了复杂的计算方法来降低这些误差，但还需要经验丰富的操作员来确保校准的正确完成。在校正时发生的错误是一些反常结果通常原因，所以我们需要花大量的时间来找出与 VNA 不同的测试模式相关联的误差来源。

8.4.2 基本测试模式和误差来源

首先让我们来做一个透射(transmission)测量(或者正向，或者反向)，参见图 8.13。通常在 VNA 之外，总是有一些电缆和夹具外接。这些外接元件引入的总电长度和损耗是变化的。更值得一提的是，这是 VNA 所不能控制的，因为夹具是用户的“权利”。

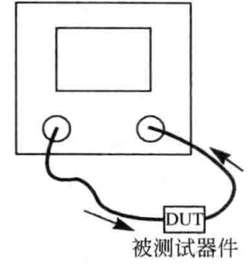


图 8.13 透射测量

一个基本的校准步骤因此是测量夹具引入的损耗和延时，然后从后续的对 DUT 测量的结果中减掉夹具的这些影响。这一步叫做通路(through，或简称 thru)测量，该步骤包括把 DUT 去除掉，然后夹具把其余部分直接连起来。然后，VNA 在给定的频率范围内测量夹具的相移与损耗，将数据存下来供后面减去用。

在通路测量之后，DUT 被加入到测试系统中，用 VNA 测量 DUT 的插入损耗和相移。在大多数情况下，我们对时延比对相位更感兴趣。因为时延就是简单的相位对频率的导数(要加负号)，VNA 很容易从相位的数据中计算得到时延。然而，要得到正确的测量结果，这里有些细节之处要注意到：其一是仪器对相位的测量是在离散的频率点上进行的，而非在整个频带连续测量。所以，这个导数只能通过近似的有限差分得到：

$$\frac{\mathrm{d}\phi}{\mathrm{d}\omega} \approx \frac{\Delta\phi}{\Delta\omega} \tag{43}$$

式(43)的分母部分是频率的间隔，叫做频率孔径，由用户控制。较小的孔径提供更加精细的分辨率，但是也可能使得到的数据对噪声更加敏感。较宽的频率孔径可以减少对噪声的敏感度，因为它是在频率间隔内的有效平均值，但也因为平均可能丢失一些细微的结构特征。现代的仪器一般都会设定对大多数情况都适用的孔径，但是用户也可以自己改变这个孔径。

另一个细节之处是 VNA 内的鉴相器只工作在有限的间隔内，即以某个相位为模。典型的鉴相器的范围是 $\pm\pi$，因而 VNA 不能区分这个范围外的和那些在范围内的相移。因此，纯时延的相位在与频率的关系图上(频率取为线性坐标)呈周期性的锯齿形状，如图 8.14 所示。用户可以引入一些物理上的判据或者其他方面的知识将不同的区域适当的

拼接起来，VNA 在根本上是缺乏所需要的这方面的信息。但是，在作图时画成这样的锯齿形也是有一个好处：对于给定的分辨率这样做可以减少垂直方向上的总高度。

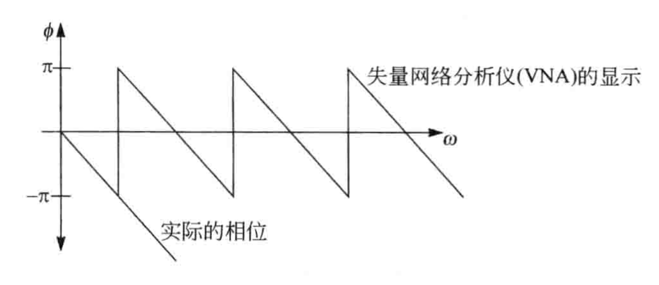


图 8.14　相移与频率的关系：理想情况与在 VNA 显示出来的情况相比

与在 ϕ 上取模有关的一个后果是：如果 VNA 在计算时延用到的频率孔径值对应的相位步长超过 π 弧度，那么显示出来的时延就是错误的。为了防止这类错误的发生，一种实用的方法是同时检查相位和时延的曲线，而非只观察时延曲线。对结果是否合理的这些简单检查可以迅速进行，因而实在是没什么借口不进行检查。

除了透射测量，VNA 的另一个基本工作模式是反射测量。跟开槽线一样，找到一个参考平面是非常关键的，我们仍然可以利用开槽线的短路方法来得到参考平面[5]。因此，作为反射测量的最简单的校准方法，这里采用了最佳的短路负载，将其连接到 DUT 所在的测量端口。利用 VNA 测量并记录在给定频率范围内的反射信号的幅度和相位，并以此为依据确定参考平面和对夹具带来的损耗进行修正(VNA 不能利用通路测量得到的关于夹具损耗的数据，因为通路测量并不给出在夹具路径长度上的损耗)。在这一步的测量之后，短路电路的 S_{11} 参数就可以画出来，如图 8.15 所示，所有的数据点应该是分布在 −1 点附近的挤在一起的点。

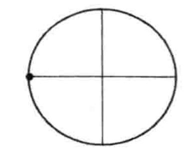


图 8.15　经过短路负载校准之后的理想化的 VNA 显示

如果没有观察到密集分布(比如观察到弧形分布)的数据点，则需要仔细地检查一下夹具(特别是连接器)，纠正任何存在的问题，然后重复进行测量。在再次验证之后，VNA 已经可以测量单端口的反射系数了。如同我们在开槽线的情形时看到的那样，这种测量实际上跟阻抗的测量是等价的。根据实际情况，用户可能希望显示的数据是反射系数或者阻抗。现代的 VNA 可以以极坐标的形式显示 Γ 或者在 Smith 圆上显示阻抗。在图 8.15 中并没有特别地标记出 Γ，因为在短路这种特殊的情况下，这些数据是在同一点上的。

一个微妙的话题是短路负载的校准同时确定了在短路处的参考平面位置，误差在校准标准范围内。夹具可能在该平面之外增加一些物理长度。我们可以通过将实际的 DUT 端短路来进行校准的方法来修正这个误差。另一个可选方法是利用现代 VNA 带有的“端口扩展”功能，就是测量仪器用算法来增加长度，因而有效地将参考平面向远离 VNA 连接器的方向移动。正确的长度增量的确定方法是：先在 DUT 端口形成尽可能好的短路，然后改变长度增量，使数据点在 Smith 圆的 −1 点附近达到最优分布为止。

⑤　原则上，将其开路也是可以的。但是，在物理近似上，短路会比开路更好。

还有一个问题是，负载为开路的情况下，VNA 的测量结果不是一个点。这是因为存在边缘效应，总会有一些容性的负载与开路的端口并联。端口还可能会发出一些辐射。如果 VNA 是被“正确地”校准了的话，测得的最终结果是一个弧形的分布。在修正这个所谓的测量问题时，一个普遍容易犯的错误是做一些想当然的对端口延伸或者时延设置的调整，这实际上会使校准的质量变差。

前面的讨论都在强调如何消除外部夹具引入的误差。通过通路校准和短路校准的方法，可以减少 VNA 在透射和反射测量时的大部分误差。VNA 可以在外部校准的辅助下识别自身的内部误差，从而得到更精确的结果。要了解 VNA 如何进行这些额外的修正，就必须要明确不同的校准所针对的误差种类。

VNA 依靠定向耦合器将信号分解为入射分量和反射分量。与其他所有东西一样，实际的定向耦合器并不完美。为了量化这些不完美性，我们需要定义各种品质因子，以应用于 VNA 中的定向耦合器(见图 8.16)。应该注意到，图 8.12 和图 8.16 都利用了简单的符号表示定向耦合器。我们清楚地知道，定向耦合器一般说来应该是一个四端口的器件，但是典型的 VNA 只利用其中的三个端口，剩下的那个端口(隔离输出)接匹配负载。

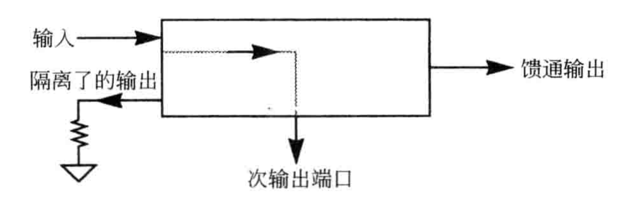


图 8.16　定向耦合器的端口定义

回顾一下表征定向耦合器的那些参数诸如耦合系数、隔离度和定向度等。特别是对于 VNA，我们希望具有无穷大的定向度，但是所有的实际耦合器都达不到这个理想的值。有限大小的定向度是误差的一个重要来源，而修正这个误差是对 VNA 进行校准的一个主要目的。

为了说明定向度误差是如何影响测量精度的，我们首先通过图 8.12 和图 8.16 来回顾一下定向耦合器在 VNA 内部是怎么连接的。注意到每个定向耦合器的主要输入都是连接到 DUT 上的。因此当进行反射系数的测量时，频率综合器(简称频综)的信号输出是用来驱动透射(through)输出这个端口。频综的信号功率逆向通过定向耦合器，到达 DUT。被 DUT 反射回来的功率注入到这个定向耦合器的主输入端口，这些被反射回来的功率一部分通过辅助端口输出，以便采样和测量。

如果定向度是无穷大的，那么传到辅助端口的能量将全部来自于 DUT 反射出来的能量，因而就能够直接测量反射能量。但是，有限大小的定向度意味着一部分从频综传输到主要输入端口的功率也泄漏到了辅助端口。所以 VNA 测量得到的辅助端口的信号是正向和反向传输功率加权和。因为它们随频率变化可能是同相或异相的，定向度不完美的典型表现是，比如说，在测得的作为频率函数的反射系数值的抖动。

下面用一个有代表性的计算来说明这个问题的严重性。假设我们有一个 10dB 的耦合器(即，耦合因子 $C=10\text{dB}$，译者注)，其定向度(D)为 30dB，我们试图测量的阻抗的回波损耗(RL)为 20dB。也就是说 $C=10\text{dB}$，$D=30\text{dB}$，$I=C+D=40\text{dB}$，$RL=20\text{dB}$(I 是隔离度，

isolation，译者注）。被 DUT 反射回来的信号功率比入射的信号功率低 $RL=20\text{dB}$，而这个反射回来的信号又经 $C=10\text{dB}$ 的衰减进入了辅助端口，因而相比于在 DUT 端的入射信号，进入辅助端口的功率比入射信号功率要低 30dB。而在辅助端口已有不希望有的相比于入射功率低 $I=40\text{dB}$ 的功率。这样我们可以看出，在这个例子中，误差功率占了有用功率的 10%，这是不能接受的。如果我们要来测量回波损耗为 30dB 的负载，情况会更加糟糕，因为误差功率就会等于希望得到的信号功率。

另一个误差来源是由界面处的反射引起的，如转接器、电缆和夹具等的界面，这个误差与定向度误差的性质基本相同。正如前面例子中提到的那样，这些误差必然会在耦合器的辅助端口产生输出，这些寄生信号可能掩盖真正从 DUT 反射回来的信号分量。

信号源的失配是另一个可能的误差来源。考虑能量从信号源流出，经过耦合器，到达负载这样一个过程。一些能量被负载反射，回到信号源。如果信号源的阻抗存在失配，那么被反射回来的信号又会再一次通过耦合器反射回去。从 DUT 反射回来的一些能量，就可能从辅助端口输出。从这个过程的定性描述来看，当负载具有很高的反射系数时，这个过程最为明显。

第三种典型的误差与我们已经考虑的问题相关——频率响应。耦合器、电缆和适配器，以及系统中实际测量幅度和相位的部分，其性能都有可能随频率的变化而变化。

定向性、信号源失配和频率响应这三种在一个端口上产生的误差，可以通过做三个实验来加以消除。例如，将一个完全匹配的负载作为 DUT。在这种情况下，定向耦合器的辅助输出端口应该没有信号。如果有信号，则说明实际存在一个定向性误差。此时，可测量并记录这个误差，之后可将其消除。这个定向性误差被消除的程度决定于在这一校准过程中使用的"完全匹配"负载的完全程度。

一般选择的另外两个实验是利用短路和开路。与完全匹配的负载一样，VNA 测量最终的精度决定于负载的阻抗接近于零和无穷的程度。在高频条件下，开路电路特别难实现，因为寄生电容很难控制。为了强调实现中可能遇到的困难，我们举个例子：一个 0.1pF 的寄生电容（这正好是 APC-7 导线开路时的值），在 10GHz 时阻抗仅为 160Ω。同时，开路端的辐射问题也随着频率变高而更加严重，这些损耗相当于产生了与寄生电容并联的阻抗的实部（即电阻，译者注）。这个问题可用如下方法缓解：将一根短路线沿导线滑动，直到短路线距离参考平面 1/4 波长时停止。这根短路线形成了一个闭合结构，可以防止辐射。

在进行两个端口的校正操作时，在这三个实验的基础上还需要增加一个：通路测量，以此来表征夹具的频率响应，如前面已描述的那样。所以，这四个实验通常被称为：短路，开路，负载，通路二端口校准方法（SOLT：short/open/load/thru）

在 SOLT 技术中，可以做一些小的改动，总的目的是为了解决阻抗标样的缺陷问题。其中的一个改动就是用一个可滑动负载来代替固定匹配负载。在这个装置中，一个接近于完全匹配的可滑动的负载沿着一根空气线（原文如此，译者注）滑动，而整个装置是用来取代固定负载的。当这个负载沿着线滑动时，那些小的反射信号与入射信号以周期性的形式相结合（交替性地相加或者相减），而得到了在复平面上呈圆形分布的数据集。定向度矢量就是这个圆的中心。由于三个点能唯一地确定一个圆，原则上只需要三次测量就可以确定这个圆。但实际上，需要测量更多的数据点来改进误差估计。

另一个改动是将滑动负载用补偿负载代替，这相当于一个不滑动的滑动负载。如果两

个点及补偿偏移的角度已知，阻抗圆的圆心就可以确定下来。而这两个点是通过两个不同长度的负载来得到的。这种滑动或补偿负载在毫米波频率上非常流行，因为在这个频率范围内，得到理想负载的好的近似根本不存在。在这些频率上，可以在紧固法兰盘之间插入一个厚度已知的薄垫来进行第二次测量。

还有一种校准方法被称作透射-反射线(thru-reflect-line，TRL)方法[⑥]。这种方法可以校准的误差跟 SOLT 方法相同，但是对用作校准负载的阻抗标样的精度要求没有 SOLT 那么高。正如这种方法的名字所隐含的一样，这个方法的第一步是将外部夹具的两个端口连接在一起，构成一个低反射率的通路，来表征夹具的特性。第二步是在两个端口分别连接一个粗略的失配电阻(这就是为什么名字里有反射这个词)。这个失配电阻的精度并不是太关键(虽然反射系数的相位应该在大概 90°以内，因为绝大多数的响应都是周期性的)，幅度也并不需要知道，只是在两个端口处的反射系数都必须要很高(通常意义上的短路经常被使用)。最后，这两个端口通过低阻夹具再次连接起来，但这次用了与第一次的透射测量长度不同的电缆(或者说是不同的夹具)。再次考虑到实际鉴相器的测量精度的局限性，这次采用的电缆线的标称长度要比在透射测量时用的线长出频带中心频率处波长的 1/4。为了得到最佳的精度，最好是将频率跨度限定在一个值，这样可以保证不同长度对应的相位差不会偏离 90°太多。这个频率跨度的确切极限是随仪器不同而改变的，而且也是所要达到的精度的函数。这个频率跨度下限的典型值是 15～25°，而上限是 155～165°。这表明对一根长度固定的导线可以校正的频率跨度大约为 8：1。如果需要更大的频率跨度范围，那么就需要多根不同长度的导线了。

TRL 方法对非同轴线系统，如微带线这样的阻抗标样很难实现(或者市场上根本就没有)的系统特别适用。对于同轴电缆线，TRL 的方法也是有优势的，因为这个系统的阻抗标样非常昂贵，而 TRL 过程并不需要昂贵的元件(至少在原则上是如此)。

还有一种方法，可校准的误差与 TRL 相同，但是它用在片上测量上更加容易，被称作导线-反射-匹配(LRM：line-reflect-match)，这个方法不需要使用不同长度的线[⑦]。这种考虑在片上测量时特别的重要，因为此时到处移动探针进行校正非常不方便。除了不方便之外，电测量的可重复性也受到接触的可重复性的限制。到处移动探针显然对可重复性不利。而且在许多情况中，根本不能够改变探针到探针之间的距离，这也使得 TRL 校准方法无法使用。

为了得到像 TRL 那样的校准结果，而又不要求有长度可变，或者是太长的校准标样，于是 LRM 方法利用了良好的宽带匹配(LRM 中的 M)。鉴于此，校准的最终效果取决于这个匹配负载的品质。这种测量方法一个特别的弱点是当用探针接触任何一个结构时，通常都会遇到一个串联电感。因此，实际中 LRM 校准的目标都是要测量出这个电感，然后再通过算法将其消除。

---

⑥ 这个校准装置是由 G. Engen 与 C. Hoer 在下面这篇文章中首先提出来的："Thru-Reflect-Line：An Improved Technique for Calibrating the Dual Six-Port Automatic Network Analyzer"(穿通-反射线：一个用于校正双六端口自动网络分析仪的改进技术)，*IEEE Trans. Microwave Theory and Tech.*，v. 27，December 1979，pp. 987-993。

⑦ D. F. Williams and R. B. Marks，"LRM probe-tip calibrations using nonideal standards"(采用非理想标样的 LRM 探针端校准方法)，*IEEE Trans. Microwave Theory and Tech.*，v. 43，February 1995，pp. 466-469。这并不是描述 LRM 的第一篇文章，但它很好地总结了与使用 LRM 有关的一些实际的考量。

正如这个方法的名字所隐含的，LMR 方法的其他两个标样是一根线与一个高反射系数的负载。后者通常就是一个开路电路，因此 LRM 标样是那些用在 SOLT 校准的标样中的一个子集。对宽带匹配的依赖实际上意味着必须依赖由专业厂家提供的标样，如 Cascade Microtech 公司。这就像在采用其他方法时，我们通常需要购买校准标样一样。

8.4.3 对微带线的一些专门考虑

就像前面的一节中所提到的那样，微带线的情形为校正造成了一些挑战。举例来说，考虑测量一个晶体管 S 参数的微带线装置。对于这样的非同轴结构，进行诸如 SOLT 的校准过程并不是那么完全明显的。既然校准的一个重要目的是去除夹具对测量带来的影响，我们显然希望在不同阶段的校准过程中去实现及应用短路、开路、匹配负载和直接通路，所有这些都在 DUT(此时是一个晶体管)所在的位置上。开路听起来是够简单的(其实不是这样，因为总有边缘杂散电容存在)，直接通路听起来也很简单。开路只需要不加入 DUT 即可，而直接通路只需要找到另一个相同的夹具，而在其中有一条直通输入与输出端口的微带线相连接。实现良好的短路近似也是直截了当的，需要第三个夹具(仍然在其他方面与第一个相同)，并将这些端口与地短路(例如，通过一片短且宽的铜箔)。最困难的是实现良好的匹配终断。例如，在微带线的终端加一个表面贴片电阻，可以作为一个粗略的原型，但是对于精确的表征，这样做并不能满足要求，因为贴片电阻的串联电感和并联电容会引起这个已经“匹配”了的负载阻抗随频率变化。

当我们重新考察一下用一个匹配的负载与 VNA 相连接时哪类误差是可以被校准的，就能找到解决这个问题的合理方法。在执行 SOLT 校准的这一步时，VNA 内部的定向性误差大部分可以被消除，因此完全没有必要再考虑夹具的其他部分的影响。所以，普通的同轴匹配阻抗标样可以直接连接到 VNA 的端口，而根本不需要担心是否以后会用到微带线。我们将这种方法称作改进的 SOLT 技术。

遗憾的是这种改进的 SLOT 技术不能校准 APC(Amphenol Precision Connector，一种精密同轴连接件，译者注)端口之外的失配而引入的在有效定向性上的误差。所以，如果过渡部分性能不好或者需要得到精确的结果，那么可以使用 TRL 校正方法。因此要完成完美的校准，TRL 校正方法是最佳的，但是改进的 SOLT 方法通常依然是一个足够好的折中选择。

最后需要考虑的是采用微带线实现开路条件是非理想的(这与所有其他开路结构都一样)，这是由于存在边缘杂散电容(数量级在 50fF)。对这个电路的部分校正是用软件实现的连接器减除。很多 VNA 都具备这样的功能，依靠软件来消除连接器的影响。可惜的是，微带线并不是这些通常提供的选项之一。最接近的选项或许就是 APC 了，它的约为 100fF 边缘杂散电容值与在 FR4 基板上的典型开路线的值非常接近。但是，仍然存在着残余误差，我们可以应用许多 VNA 具备的“可变端口扩展”的功能来大幅减小这些残余误差。

最后，我们总是要遵守测量仪器的功率限制，以避免工作在非线性区，甚至造成损坏。一般说来，加在仪器上的功率要小于 20dBm(100mW)以避免损坏，而采用低于 0dBm 的功率以避免仪器出现非线性。这只是一些粗略的准则，所以在使用时需要参考具体仪器的使用说明以得到正确的测量值。

8.5　校准方法总结

前面这一节列举了太多的校准方法，以至于很容易搞糊涂（如果我们涉及了所有存在的方法，就会更搞不清楚了）。下面是一个对校准方法的总结和评论，来帮助读者回顾这些方法的优缺点，以便对这些校准方法做出合理的选择。

最简单的校准方法是对单端口或二端口的短路通路校准，但它们只能校准外部夹具和检测器的频率响应误差。由于有限的定向性和信号源失配而引起的误差是不能校准的。这种方法也被称作响应校准。

对于单端口的测量，更好的校准方法是短路-开路-负载组合（short-open-load，SOL）。只要阻抗标样是准确的，这种方法就可以消除来自于有限的定向性、信号源阻抗失配和检波器频率响应带来的误差。再加入一个直接通路测量，就构成了 SOLT 校准。SOLT 是一种二端口的校准方法，该方法可以校准 SOL 所能校准的所有误差，还能校准二端口夹具引入的额外电缆的误差。

基本的 SOLT 方法可以有很多变种，包括用滑动负载或者固定阻值的补偿负载来代替该方法中的负载部分测量。改进的 SOLT 方法将 SOT 测量与微带线（或者其他非同轴）夹具结合起来以消除 VNA 固有的定向性误差之外的其他误差。通过消除（或基本消除）定向性误差，采用没有夹具的标准同轴匹配负载进行的校准就完成了整个校准过程。

直通-反射-导线（TRL）这个校准方法规避了以优质阻抗标样作为校准负载的需要，同时可校准与 SOLT 方法相同的误差。TRL 方法尤其适用于带有微带线那样的非同轴系统。采用 TRL 校准，在 18 GHz 下可以将定向性误差和信号源阻抗失配误差减小到－60 dB，同时也基本上消除了频率响应误差。这些值可以与采用典型的 SOLT 方法（固定负载）达到的－40 dB 定向性误差和－35 dB 信号源失配误差相比。采用滑动负载和补偿负载可使 SOLT 方法的精度提高到固定 SOLT 方法和 TRL 方法之间。

最后，LRM 校准技术得到的校准误差与 TRL 方法相同，但是 LRM 的一个优点是不需要使用不同长度的线。所以不需要改变探针的距离，降低了夹具的复杂度。同时，可重复性也提高了，因为在整个测量的过程中，夹具带来的误差变得更加的恒定。LRM 的缺点是需要购买只能从外部供应商处得到的昂贵而脆弱的校准基片。

8.6　VNA 的其他一些测量能力

由于大量采用计算，现代的 VNA 具有除测量 S 参数以外的其他功能。比如，在一个宽频率范围内测量了 S 参数后，得到的频率响应数据可以转化为时域响应的数据。可以从 VNA 的数据中得到阶跃响应和 TDR 轨迹。虽然用于所有测量和计算的时间要比进行“真正”的 TDR 测量的时间长得多，但是这些额外的功能仍然是很受欢迎的。

鉴于变换算法的本质，有可能通过一些数学上的小技巧来执行在实际的时域测量不能实现的功能，例如考虑当 TDR 的轨迹包含了多个反射的情况。在真正的 TDR 轨迹中，前面的不连续将掩盖后面不连续的影响。而 VNA 可以消除第一个不连续，使得进一步检测后面被掩盖的不连续成为可能。VNA 的这种对首个不连续影响的消除（叫做门控阻抗，或

者门控 TDR 测量)程度取决于 S 参数测量中的精度和噪声。

8.7 参考文献

除了在上面的脚注中引用的文献外，读者同样可以参考惠普公司的讲义《高频网络矢量测量》(Hewlett-Packard High Frequency Vector Measurement Seminar Notes, April 1989)。这本讲义中有关于 VNA 使用方法的很好的总结，还对误差来源和校准方法做了简要的讨论。另一个有用的参考文献是几乎任一个 VNA 的用户手册，比如 HP8720C(可工作于 130MHz～20GHz 的频率范围)或者 HP8510C(可工作于 50MHz～100GHz 的频率范围)。

8.8 附录 A 其他一些阻抗测量的装置

8.8.1 SWR 计

驻波比(standing-wave-ration,SWR)计是快速估算阻抗失配的经典仪器。这种仪器的最简单的结构形式是将定向耦合器(弱耦合)与测量幅度的方法结合起来。将需要测量的器件连到耦合器主线的直通端口上，如图 8.17 所示。

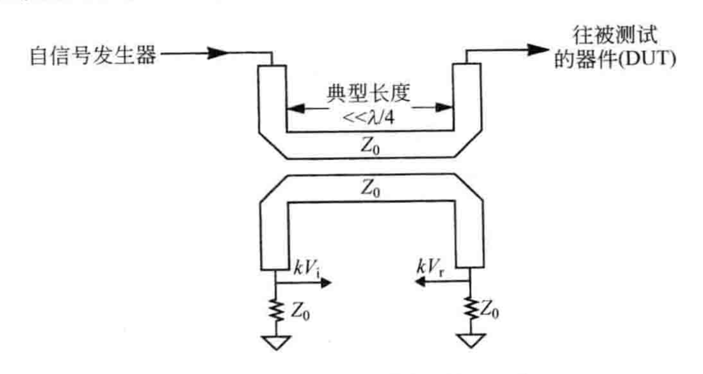


图 8.17 经典的 SWR(驻波比)计

SWR 计的最廉价的实现方式是采用基于二极管的峰值检测器和测量仪，它可以提供相对功率的指示。首先将检测器连接来测量对应于入射功率的耦合信号。一般会通过调节一个电位计使得测量仪指示的读数达到满量程。然后，将检测器连接到另一个被耦合的端口去测量被反射的功率。显示的结果经过校准直接给出 SWR 值。

很明显，其他的定向性耦合器也可以用，特别是考虑到微带线作为耦合线的耦合器有定向性差的缺点。图 8.17 中所示的边缘耦合方案只是用作说明的(当然这很便利业余实验者在一个周末就可以将其实现)。

8.8.2 栅陷式(Grid-Dip)振荡器(GDO)

在过去，业余无线电爱好者常用的一种仪器是栅陷式振荡器(grid-dip oscillator,

GDO)。虽然它并不常用于千兆赫兹频段(由厂家生产的 GDO 最高频率大约处于 UHF 频段的下限),但是并没有什么根本的原因来限制它不能用在这个频率范围,只要物理上是可实现的(我们甚至可以想象将一个这样的振荡器集成到一个探针台的探针上去)。不管怎么说,GDO 给我们提供了一个将 RF 原理和微波原理结合起来的机会。

经典 GDO 的设计思想是十分简单的。如果将一个网络放在振荡器的谐振槽路旁边,由于与外部网络的耦合,振荡器的能量损耗会增加。如果外部的网络与振荡器的谐振槽路在相同的频率上发生共振,这种能量损耗将急剧增加。此时,向外界负载提供的这个能量使得振荡器的偏置电压或电流降低(在用真空管实现的时代,这反映在栅电极的偏置电流上,这也是 grid-栅-这个字出现在 GDO 名称上的原因,译者注),这个偏置电压或电流的下跌通过一个简单的电压/电流计就能测得。所以,GDO 由一个可在某频率范围内调谐的振荡器,一个允许与外部网络耦合的探针,以及一个指示电压/电流下跌的电压/电流计组成。

GDO 的用途广泛。在精确校准后,通过 GDO 很容易确定外部腔体的谐振频率,这样就可对外部电路进行粗略的调频(精度大概在几个百分点范围内)。如果将一个未知的电感连接到一个已知大小的电容上,通过 GDO 测得的谐振频率可以计算未知电感的大小。同样,用一个已知大小的电感,也可以测量未知电容的大小。

通过一个简单的改动,就可将 GDO 用于频率测量,这是将其兼用作本地振荡器和混频器来实现的。将 GDO 与外接的振荡器耦合起来,可以起到混频的效果,这是由真空管本身带有的非线性引起的。直接将耳机接到 GDO 振荡管的极板电路中,可以听到拍频随着 GDO 振荡频率的不同而变化。当听得到的击打频率变为 0 时,GDO 的频率就被调整到外接振荡器的频率上了,这个值可以直接从校准过的读数上得到。频率大约 1%的确定精度是可以达到的,这主要取决于读数校准的质量和 GDO 的稳定性。

图 8.18 是将一个典型的 GDO 用 JFET 实现的电路图[⑧]。商用的 GDO 会同时提供一组精确的插入式螺旋线圈,这些线圈的电感值都具有优于 1%的精度。用户从中选择一个电感,其值对应的 GDO 频率跨度应该差不多位于所要求的频率范围中间。插入式线圈不仅仅是用作 GDO 谐振腔的一个元件,同时也用作连接到 DUT 的探针。

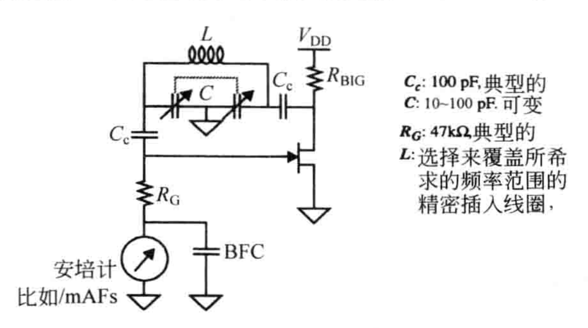


图 8.18 基于 JFET(结型场效应管)的“栅陷式”振荡器(GDO)

⑧ 过去的几十年中,在数不清的出版物、电路套装和商用仪器中,有各种各样的基于这个电路的版本。我们这里提供的实现主要是基于《无线电业余爱好手册》(*The Radio Amateur's Handbook, 29th ed., Rumford Press, Concord, NH*, 1952)上的一个电路。

为了正确地使用 GDO，特别要注意避免将该装置和 DUT“过耦合”。这是因为耦合会改变谐振的模频率。所以，我们在耦合 GDO 和 DUT 时，耦合强度应该适中，这样才能得到可靠的压降。典型的做法是，使用者最开始将装置与 DUT 靠近，此时就可能出现过耦合，使得指示计的读数偏大；随后，当将装置逐渐远离 DUT 时，不断地调整 GDO，以追踪下跌的轨迹。

8.9 附录 B 设计课题

8.9.1 微带槽线设计方案

如我们已经看到的那样，现代的网络分析仪确实是功能非常强的仪器，它能够非常准确地表征微波网络的宽频带特性。遗憾的是，实现这种功能是付出了很大代价的：一个典型的吉赫兹 VNA 的价格比运动型多功能车（sport utility vehicle，SUV）的价格还贵，在二手市场中，也鲜有大幅度降价。显然，这样的 VNA 价格对大多数业余爱好者来说都承受不了（就是很多搞学术研究的实验室也不能承受）。所以，对于那些缺乏经费的用户，槽线（slotted line）结构是另一种可行的选择。除了价廉之外，槽线是讲解史密斯图原理的很好的教学工具（例如，可以给出“朝向信号发生器的波长”等术语的清楚的解释）。如前面提到的一样，许多槽线在二手市场上能以十分合理的价格购得，起码对那些用在吉赫兹频率范围低端的槽线是如此。

这一节我们介绍一种便宜得多的（也简陋得多的）替代装置——微带槽线系统——可以用来测量阻抗和频率（5GHz 或更高）。这个仪器（这样的叫法是有点夸张）具有一些重要的特点：非常便宜（总的元器件和材料费用不会超过 5～10 美元），而且用普通的工具和材料就可以制作出来。缺点是仪器的精度并不是特别的好，而且简陋的机械装置并不很鲁棒。但是，它的性能对于吉赫兹频率范围低端的业余应用是完全够了。在制作中再多加小心，这个装置的性能完全可以满足大多数的学生实验项目的应用要求。

仍然是出于降低成本的考虑，在这里描述的设计是基于 FR4 的电路板材料和成直角安装的 BNC（Bayonet Neill-Concelman，译者注）同轴电路连接件，这种连接件在有关微带线的章节中已经描述过。特别是制造微带线比制造开槽同轴空气电缆要容易得多。当然，应用低耗的 PC 板材料及性能更好的连接器可以得到更好的性能。读者当然可以在预算、耐性和性能要求的前提下作出各自的变动。

第一步是准备一段比需要测量的最长的电波长更长的 FR4 板材——但是不要过长以至于在预定的工作频率范围外额外损耗严重。当最小工作频率为 1GHz 时，较好的折中是 25cm。一条这么长的导线，在 1GHz 时，损耗为 0.8dB，在 5GHz 大概为 4dB（与此相比，实际共轴槽线的损耗在最坏情况下是小于 1dB 的）。如果仅将仪器应用在更高的频率，那么将这根线变短可以减少损耗，从而改善性能（我们只需要这根线的长度能包含几个最小值就可以了。注意，经过每个波长的损耗是恒定的，其值为稍小于 1dB 每波长）。

在 FR4 板的两端装上 BNC 连接件，然后用铜箔带构建一条 50Ω 的微带线。铜箔应该尽可能地铺平。下一步是，将不导电的公制米尺贴在微带线的下面（如果你没有合适的尺子，用图 8.19 放大两倍的复印件替代）。由于复印的精度可能有相当大程度的变化，需要检

查直尺的标度在被放大时,有无变形。在 5mm 范围内(一个标度,译者注)仔细地插值读数,应该可以达到约 1mm 的精度,虽然准确度是另一回事。

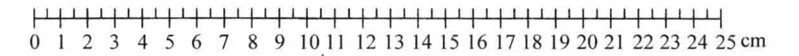


图 8.19　微带线的公制标尺(所画的尺寸为实际的一半)

下一步是构造一个检测器。这里我们用基于肖特基二极管的检波电路,这个电路是通过电容耦合到微带线上。这是一个简单的电路,而面临的最大挑战是在机械上的:如何在构造一个游标的同时保证检波器与微带线的恰当且一致的耦合。

探针就是一根普通的针(就像新衬衣上带有的那种),仔细地将其插入游标一端的底部,然后用一小点环氧树脂固定,剪切到合适的长度(图 8.20 所示)。这个探针用一个较短的绝缘材料(最好是从连接线上取下的特氟纶套管)包住。它们相当于位于微带线和探针中间的电介质,同时也提供了平滑的滑动效果。探针和线的移动都需要足够的平滑,因为这样才能使探针沿线滑动时候的耦合恒定。

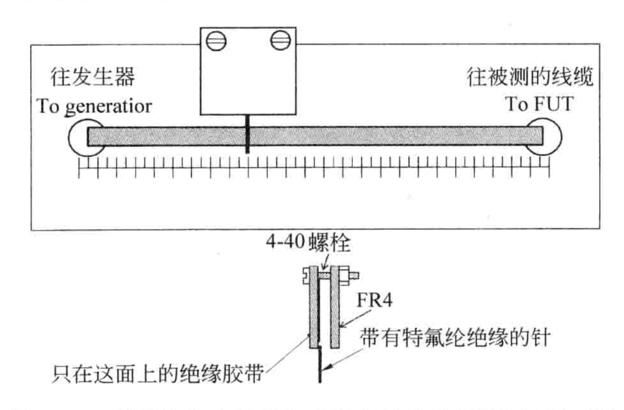


图 8.20　开槽线的底视图与探针组件的侧视图(不成比例)

这个滑动组装件是由两条用螺栓固定在一起的 FR4 组成。特氟纶带(或者非常光滑的铜箔带基带)可以被粘贴在那条粘有微带线 FR4 板条的内侧面,以减小滑动摩擦和磨损。它的厚度也需要非常精细的控制,以确保探针和线的接触良好。

两块 FR4 有铜箔的面都朝向主板。与主板的接地面相接触的铜箔为探针电路提供了接地点。这个游标包含了探针电路,它由一个二极管检波器和一个电阻性负载构成(见图 8.21)。

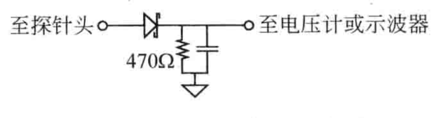


图 8.21　探针组件的示意图

这个肖特基二极管只要是低电容,高频的器件(如 HP5082-2835,或者-2860)都可以用,其阳极端口与实际的探针相连。输入信号的幅度应该足够小以使得二极管大致工作在平方律的特性区,在这个区域内输出电压大致与功率成正比。这个正比常数,只要它保持为一个常数,大小是无关紧要的,因为我们在计算 SWR 时,只需要计算电压的比值。这个平方律关系使得我们通过计算探针输出电压的开方就可以得到 SWR。

图 8.21 所示的电阻负载为 470Ω,但是读者可以尝试改变阻值的大小以使得检波器的工作范围达到最大。贴片电阻被用来使寄生参数最小,虽然检测器的输出端对典型大小的

寄生参数并不敏感。

游标的电容，再加上与其相连的测量计与示波器的输入电容通常是足够大了，因此并不需要再添加额外的电容。为了最大化的使用范围，我们希望可以检测出亚毫伏量级的电压，此时，就需要对放大检波器的输出进行一些放大，以使得其输出范围可方便地采用不那么昂贵的仪器进行测量。

那么做到了所有前面讨论所要求的后，微带槽线到底性能有多好呢？FR4 和直角安装的 BNC 带来的损耗，探针耦合的不规则性，缺乏对探针的校准和槽线本身是用手工组装的，等等，所有这些都将使得这个阻抗测量仪器显得相当粗糙。大致说来，对于阻抗值在 $Z_0/5$～$5Z_0$ 的范围内，我们可以期望有合理的精度。对于最常见的与 Z_0 实现良好匹配的情形，这个工具特别好用，可以轻松地得到小于－15dB 的 S_{11} 值。在 1GHz 频率下得到的结果通常好得惊人，但随着频率增大到 5 GHz 以及更高频率时，得到的结果就逐渐变差了。

如果这个装置只在更高的频率下使用，有几处需要改进的地方。采用直的 SMA(Sub-Miniature version A，译者注)连接器来代替成直角的 BNC 连接件，将 FR4 用 RO4003 代替，并且不采用手工方法制成的铜箔微带线，因为这样的微带线一致性较差，改用常规的 PC 板制造技术。最后，需要采用性能更好的二极管(例如 M/A- COM MA4E2054，它的性能指标满足 10GHz 以上的频率应用)。如果槽线的长度缩短五倍左右，并且配合上面提到的各种改进方法，就有可能在 5～10GHz 的频率范围内达到令人满意的工作状态。

采用槽线进行频率测量

如果频率测量在 10％的精度范围内是可以接受的，那么槽线是替代吉赫兹频率测量仪器的非常经济的装置。驻波的最小值之间的间隔为半个波长，计算频率只需要再知道波的传播速度。对于 1.6mm 厚的 FR4，有效介电常数为 3.5，误差在 5％～10％范围内。这样板上的波长是自由空间波长的 0.535 倍。这个值在 1GHz 时对应的波长是 160mm。那么来进行频率测量，我们只需要简单地将线开路和短路以使得 SWR 最大，同时测量最小值之间的距离。可以通过测量几个最小值之间(如果存在的话)的距离，然后再除以半波长的个数来提高测量的精度。将测得的值乘以 2 就得到了波长。用这个以毫米为单位的波长去除 160mm，就得到了以吉赫兹为单位的频率。

举一个数值的例子，假设我们测得最小值之间的距离为 25mm，则波长为 50mm，将这个值除 160mm，得到的频率为 3.2GHz。

不经任何校准，可以得到 10％左右的绝对精度。经过校准，误差可以减小 5～10 倍，这取决于构造槽线以及进行校准时的细致程度。此外，这根槽线可以测量的频率比进行阻抗测量时的工作频率要高得多，因为在频率测量时，只需要知道最小值之间的距离，而完全无需关心诸如损耗之类的特征。

8.9.2 自制亚纳秒级脉冲发生器

制造快速脉冲和阶跃信号发生器的技术非常的专业，要想借助典型的普通实验室所具备的条件来制造出上升时间可与现在最先进的仪器相媲美的脉冲发生器是不现实的。但是，我们会愉快又吃惊地发现，利用业余爱好者可以找到的一些零件，制造上升时间为 200ps 左右的脉冲并不困难。这种脉冲发生器对于评价示波器的品质，特别是示波器探针特别有用。

对这种脉冲发生器进行很小的修改就可以将其转化为一个可触发的脉冲发生器，脉冲的上升时间为 200ps。在几种产生具有这样快速脉冲的方法中，对于业余爱好者而言最经济的方法无疑是利用晶体管的一个非正常工作模式——雪崩击穿。在这种类型的击穿中，集电极电压高到足以使电子从围绕原子的轨道上被剥离出来，这样就形成了电子空穴对。电子朝着正电极（这里假设是集电极）被加速，而空穴朝着基极被加速。随着这些游离的载流子被加速，其中一些就会撞到其他的硅原子上，形成更多的电子空穴对，如此反复，使得集电极电流迅速增加。

图 8.22 中所示的简单的脉冲发生器就是利用了这种雪崩击穿。在这个电路中，集电极的电源电压大大高于晶体管的击穿电压，而其确切的值并没有什么关系。但是，在任何情况下都不能将市电的交流电源直接用作驱动电压，这样做实在是太危险了。相比之下，我们更推荐用电池来驱动的电路。另一个特别容易得到的高电压是一次性相机中的氙气闪光灯电路。这些一般可以从附近的照相馆很便宜地买到（甚至免费）。在这种电路中，大电容两端的电压为 200～300V，而这种量级的电压足以使我们感兴趣的绝大部分晶体管都置于雪崩工作状态。当将这个电路板从相机里面取出时，必须十分小心，当然，用导线进行连接时也同样要小心。这个大电容可将大量电荷存储相当长的时间，而如此大量的电荷是相当危险的。

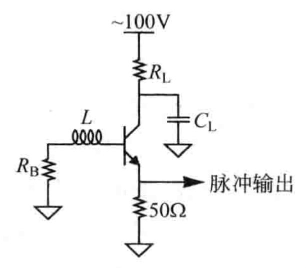


图 8.22　自激雪崩模式脉冲产生器

电路图中的电容 C_L 可以通过将铜箔带置于 1.6mm FR4 板材上的接地平面之上构成。每边大约为 0.75cm 的方形条是良好的起始值。可以通过增加或减少铜箔来调节脉冲的宽度和幅度。减短接到集电极上的导线长度非常重要。在集电极和 C_L 之间的任何电感都可能会减慢脉冲。只需要几个纳亨量级的电感就可以使电路减慢两倍以上。

脉冲的重复率是由负载电阻 R_L 与 C_L 一起决定的。重复频率的一个合适的目标约为 100kHz 附近，对应的 R_L 为 100kΩ 到几兆欧之间。重复频率低通常伴随着抖动的恶化，而重复率太高又会导致功耗过高。

一旦雪崩击穿开始，集电极的电压急剧增加有两个原因：在集电极区雪崩电子倍增的直接效应和由于雪崩产生的空穴使得基极电流的增加。这个增大了的基极电流，通过晶体管正常的工作机理使得集电极的电流增大。这个正反馈的机制通过提供相对较大的基极负载得以增强：由基极流出的空穴电流通过这个电阻使基极电压大为增加。R_B 的典型值在千欧到几十千欧之间。确切的电阻值并不重要，但是如果要优化某个给定的电路，可以从调节这个电阻值开始。类似的，串联寄生电感并不是个问题（有时候，甚至是有利的）。为了强调这个电感的用处，图 8.22 的电路特别将其画出。

脉冲的宽度取决于集电极电容的大小（越大的电容可产生更高更宽的脉冲）和晶体管的特性。较小的集电极-基极电容有利于基极和集电极的电压在正反两个方向迅速来回变化。一个更为关键的参数是 BV_{CBO} 和 BV_{CEO} 之间的比值。前者是在发射极开路时测得的集电极-基极击穿电压的量度，后者是基极开路时测得的集电极-发射极击穿电压。后者一般比前者小，这是由前面提到的内部正反馈机制引起的。对于大多数小信号晶体管，这两个电压的比在 1.5～2.0 左右，但是对一些晶体管（如 2N2369 或者 Zetex FMMT-417）这个比率超会过 2.5。这种晶体管特别适合制作雪崩脉冲发生器。

需要强调的是，晶体管生产商几乎从来不标出哪些晶体管适合在雪崩击穿模式下工作。即使我们找到一个能在雪崩击穿下工作的晶体管，也并不意味着我们能指望同一类型的晶体管都会很好地在雪崩击穿状态下工作。因此，一般说来需要经过一些手工的筛选。即使是这么说，对于一批 2N2369 晶体管中，典型的是有 75%以上可以在雪崩模式下很好地工作，它们对 50Ω 的负载可以提供峰值在 5～10V 之间的脉冲，其上升下降时间接近 200ps。这样的速度是 60 年代昂贵的实验室设备才能达到的值。1A/ns 左右的电流压摆率（slew rate）很难通过常规方法得到，所以需要筛选一些晶体管的这个代价并不算大。这么高的压摆率也强调了减少发射极电路串联电感的重要性，要知道一个纳亨量级的杂散电感就会使电压降低整整 1V！

脉冲发生器在高速电路中用途广泛。作为一个具体的例子，通过观察示波器显示的上升和下降时间，采用这样的脉冲产生器可以迅速得到示波器-探针组合的带宽。仪器缺陷或对仪器和电缆的校准不当引起的测量结果异常也是容易观测到的。考虑到常用的替代办法是通过在吉赫兹范围内通过正弦信号发生器扫描来测量频率响应，采用脉冲发生器明显是一种相当廉价的选择。

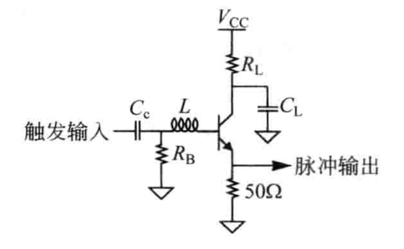


图 8.23　触发激励的雪崩模式脉冲发生器

通过简单的改动就可以将自激脉冲发生器转变成触发激励器件，见图 8.23。这里，电源电压被调节到比自激状态稍低的一个值。然后，触发脉冲通过一个小容量（比如说，1～10pF，取决于触发脉冲的幅度和上升时间）、低寄生感抗的电容耦合到基极电路上去。这个触发脉冲会将晶体管推入雪崩击穿工作区。

第9章 微波二极管

9.1 引言

现代微波二极管的出现要溯源至在第二次世界大战期间军用雷达的需求。在那个时代的真空管无法工作在几兆赫兹的雷达频带上。幸运的是早在1930年中期的时候，贝尔实验室的科学家 George C. Southworth 在他的早期有关圆柱波导的研究时就埋下了一颗导致以后突破的种子。俗话说，需求是发明的母亲。受到他的同事 Russell Ohl 的启发，他试用晶体检波器(在那时几乎是过时了)作为功率传感器，希望这个与点接触相关联的小的电容能够允许在他正在使用的高频上(在1GHz的倍频程之内)[①]。在硅上的测试得到的令人有希望的结果证明，这种二极管确实可以在真空管不能胜任的场合成功地工作。到了“二战”结束时，MIT 辐射实验室和其他的一些实验室进行的一个集中开发项目成功地制造出能够工作在频率超过30GHz的可靠的点接触硅二极管[②]。

在这一章，我们将仔细研究这种二极管。然而，我们也将“二极管”这个词扩展到包括许多其他的用在微波电路中的二端半导体器件。这样，除了介绍常规的结型二极管和肖特基二极管之外，我们还将介绍变容二极管(参量二极管)、隧道二极管(包括反向模式二极管)、PIN 二极管、噪声二极管、急变二极管(阶跃恢复二极管)、Gunn(耿氏)二极管、MIM 二极管和 IMPATT 二极管。因为篇幅的限制，我们不得不省略了一些较少被广泛使用的但仍然非常有意思的二极管类型(如 TRAPATT，LSA 等)[③]。

这一章同时包含了一个附录(9.13节)描述如何利用家庭烤箱烘焙硬币产生的变化来制造肖特基二极管的方法，以及如何将这种二极管用在一个工作频率为1MHz微波频段的晶体管收音机。

9.2 结型二极管

二极管中被应用最广泛的还是最普通而又古老的结型二极管。大多数器件均未在物理教科书中提到，结型二极管其实在其被发明之前就已经被发现了。第一个这种二极管是在贝尔实验室偶然地做出来的。这是在进行硅晶体生长实验时，无意识地将晶锭的两端掺杂

---

① 在 Ohl 的建议下，Southworth 从当地的一个旧货市场得到几个老式的硅触须检测器，经过过几次试验和改进，最后享受到成功的喜悦。

② Torrey 和 Whitmer 在 MIT 辐射实验室系列出版物第15卷中描述了这个非同寻常的项目。(volume 15 of MIT Radiation Laboratory Series, McGraw- Hill, New York, 1948)

③ 对于其他种类二极管的综合性回顾，参看 Sze(施敏，译者注)的 *The Physics of Semiconductor Device*(半导体器件物理)(Wiley, New York, 1981)，以及他编辑的另一本书 *High-Speed Semiconductor Device*(高速半导体器件)(Wiley, New York, 1990)

了不同极性的杂质。在1940年年初，Russell Ohl 仔细测试了从那块晶锭的某个特定部位切割下来的晶片，而那个部位恰好跨越了从n型到p型的过渡。在尝试着弄清楚电阻测量不稳定的原因时，他首先发现了其光伏特性。他的太阳能电池产生的电压为0.5V，这让同事们非常吃惊，因为当时常用的亚铜氧化物和硒制作的太阳能电池所能产生的电压比这个值要小一个数量级。虽然结型二极管的整流特性在当时也被注意到了，但是点接触器件更好的频率响应性能，使后者在整个战争期间成为关注的焦点。因而结型二极管的主要发展直到20世纪40年代末期才发生，这是因为当时对点接触器件的机械不稳定性的不满意，使得工程师们更加注意到结型器件。

与点接触二极管一样，结型二极管的电流电压成指数关系：

$$i_D = I_S \exp\left(\frac{v_D}{nV_T}\right) \tag{1}$$

其中 I_S 是一个随温度变化的电流，V_T 是热电压(kT/q)，n 是一个理想因子(理想情况下为1)。对于大多数实际的二极管，这个理想因子通常是在1～2之间[④]。在室温下对正向偏置的二极管，电压每增加 $18n$ mV，电流翻倍；电压每增加 $60n$ mV，电流增加10倍。

在无线电频率下，我们必须要考虑寄生电容和结电容，以及寄生电阻和电感的影响。在反偏下，小信号电容约等于：

$$C_j = \frac{C_{j0}}{(1-V_j/\psi_0)^m} \tag{2}$$

其中 C_{j0} 是零偏置下的电容，V_j 是结电压(正偏时该值为正)，ψ_0 是名义上的接触电势(与材料的禁带宽度有关，对于硅，大约为1V左右)。m 是一个常数，它的值取决于掺杂分布。对于线性缓变结，m 等于1/3；对于突变结，m 等于1/2；如果是超级突变结(其细节将在9.4节详细介绍)，m 可能大于1/2。

真实的二极管的结电容在反偏和弱正偏的情况下都很好地符合这个公式。在强正偏的情况下，电容的典型极限值为 C_{j0} 的2～3倍，而不是像公式中描述的当外加电压接近 ψ_0 时，电容的值会变到无穷大。

这个随电压变化的电容既有好处又有坏处。当需要一个电可控的电容时，比如说在变容管的情况(更详细的介绍见9.4节)，我们希望增强这个特性。但是在其他情况，这个电容的可变性就成了不希望的失真的来源。

对二极管的一个基本要求是其正偏情况下的阻抗远小于反偏情况下的阻抗。正偏时，在高频下，其阻抗主要由寄生的串联电阻决定。在关断方向，二极管的电阻约等于电阻和结电容的串联。如图9.1所示，在这个模型中，R_j 代表基本的依赖于偏置的二极管电阻，其值在正偏条件下为

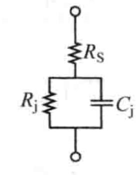


图9.1 简单的小信号二极管模型

$$R_j = \frac{nV_T}{I_{BIAS}} \tag{3}$$

其中 nV_T 与式(1)同，I_{BIAS} 是流经二极管的偏置电流。室温下，I_{BIAS} 为1mA，对应于理想因子在1～2之间，R_j 的值为25～50Ω。反偏时，我们通常假设在射频频率下，R_j 比 C_j 的电抗

④ 这个理想因子只有当表面漏电和在耗尽层复合可以被忽略时才趋近于单位值(即1)。

大很多，可以看成是无限大。

反向与正向的阻抗比为

$$\frac{Z_R}{Z_F} \approx \frac{R_S + \frac{1}{j\omega C_j}}{R_S} = 1 + \frac{1}{j\omega R_S C_j} \tag{4}$$

$R_S C_j$ 的倒数是二极管的截止频率。使得反向与正向的阻抗比最大化，等价于使截止频率比设定的工作频率要高出很多。对于标准的结型二极管，串联电阻和二极管电容与截面积的关系是相反的，所以它们的乘积与面积一阶无关。这个不随面积变化的特点也就束缚了结型二极管的工作频率范围。

另一个需要考虑的情形是，结型二极管不能突然关断。因为作为一个少子器件，只有当存储的少子电荷或者被主动地移走，或者通过复合自己耗尽，才会关断。二极管保持开启的这个时间间隔被称为反向恢复间隔，参见图 9.2。t_{rr}值的测量是从电流改变极性时开始算起，直到其跌落到某一个设定的值 I_{rr}。由此可见 t_{rr}的长短决定于设定的 I_{rr}的大小。t_{rr}也是一个正向电流的函数，随着 I_F 的增加而增长。另外，它是反向电流峰值 I_{RM}的函数，随 I_{RM}的增大而减小。

有些二极管设计成产生一个极其快的在反向恢复的末期到非导通的过渡。这个从开启到关断的急速变化会产生施加的基本信号的高频谐波。我们将在 9.8 节进一步讨论这种急变二极管（或者阶跃恢复二极管）的使用。

一种被称为正向恢复的现象很少被讨论。就如二极管不能在瞬间被关断，同理，也不可能被瞬时开启。因此，二极管的正向电压在达到一个稳态值，比如 0.7V 之前会呈现很大的过冲。不幸的是，正向恢复时间很少在器件数据手册中给出。代之是，有关正向恢复的数据通常是以在某个设定的测试条件下得到的最大正向压降 V_{FM}的形式给出。图 9.3 给出了一个简单电路的例子。通过调节阶跃信号的幅度来产生某个给定的稳态电流，V_{FM}然后就测量到了。V_{FM}比稳态的正向压降大 3～4 倍是很正常的。标称相同的二极管的正向恢复特性可能会有很大的不同。

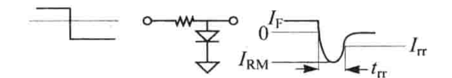


图 9.2　用于反向恢复的有代表性的测试电路与典型的波形

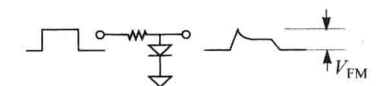


图 9.3　用于正向恢复的有代表性的测试电路与典型的波形

除了已经介绍的寄生串联电阻外，封装了的二极管通常还有其他的寄生效应，起码会存在一定量的串联电感和并联电容。图 9.4 描述的是封装后的二极管的一个合适的高频电路模型。这个模型是十分典型的，可以应用于几乎所有的封装二极管上。

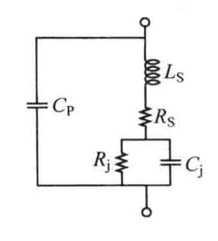


图 9.4　封装好的二极管的简单高频小信号模型

串联电感 L_S 用来表征由二极管的物理长度引入的寄生电感，而并联电容 C_P 表示的是二极管的两端间不可避免的电容耦合。当然，封装寄生参数是由封装的特点决定的。尽管这么说，对于目前在微波电路中通常使用的封装，串联寄生电感的值在 1nH 上下的两倍因子以内，而并联电容的值在 100fF 上下

的两倍因子以内。作为一个校正参考，低频二极管（如经常用到的 1N914，1N4148）通常使用 4mm 长的 PP 玻璃封装引入 3nH 的电感（假设在玻璃封装的端口上没有额外的引线）和 150fF 的并联电容。

更加复杂的模型可能在图 9.4 的基础上再添加额外的串联电感，来计及除了器件封装以外的其他因素引入的电感（如引线电感等）。

9.3 肖特基二极管

肖特基二极管（也被称为热载流子二极管）其实比各种结型二极管出现得更早。示波器雏形的发明者 Ferdinand Braun 在 1874 年发表了第一篇关于晶体整流器的文章⑤。整整 25 年后，为了改进初期的无线电技术，他回到这个先前的工作，但是他没有申请任何专利。这个专利的拥有属于有才干的 Jagadish Chandra Bose，他当时需要为他用专门的火花隙装置产生的 60GHz 辐射找到检波器⑥。他在 1901 年申请了一个有关基于方铅矿的点接触二极管用作测辐射热仪而不是整流器的专利（专利在 1904 年被准予）。在后面的几十年，这个二极管的工作原理并没有完全弄清，但这并不影响研究者们在 20 世纪的早期申请了无数的与检波器有关的专利。其中一个意义重大的专利是在 1906 年由 Greenleaf Whittier Pickard 申请的整流硅检波器⑦。

Pickard 的硅检波器对激励现代半导体的发展有着重要的意义，因为正如前面提到过的那样，就是这种检波器是被 Southworth 几乎绝望地在实验中被试过。正如他期望的那样，点接触真正保证了极小的器件的电容，远比真空管的电容小很多。更重要的是，罕见地这次墨菲法则（Murphy's law）失灵了：与点接触相关联的电阻值不是与面积成反比，而是与接触的半径成反比，即

$$R_S = 1/4\sigma r \tag{5}$$

其中，σ 是电导率，r 是接触半径。这个扩展电阻公式的适用范围为：只要点接触半径比（a）半导体的厚度（b）其他接触的半径小很多。需要注意的是：寄生电容与面积成正比。所以当接触半径缩小时，电容减小的速度比电阻增加的速度要快，因而寄生参数的 RC 乘积随半径线性减小，而不再是保持一个恒定的值了。这个缩放行为与平面结型器件有很大的不同，这也是点接触器件有很好的频率响应的原因。

因此，要得到好的高频响应特性，关键是在保证形成良好接触的前提下尽可能地减小接触面积。在制造非常锐利的针尖方法（通常是用钨）以及制造出很薄的打磨得很光的硅片技术的进展，使得人们在 20 世纪 40 年代早期就可以制作出能在 K 带频段工作的点接触肖特基二极管⑧。需要提到的是，大致 2μm 的接触直径直到 20 世纪 80 年代才能通过集成电路光刻实现。

---

⑤ "Ueber die Stromleitung durch Schwefelmetalle," *Poggendorf's Annalen der Physik und Chemie*, v. 153 (now v. 229), pp. 556-563. 为了表彰他的很多重要贡献，他与 Guglielmo Marcon（马可尼）共同获得 1909 年的诺贝尔奖。

⑥ 这不是印刷错误，Bose 在 19 世纪就开始了毫米波辐射的研究。

⑦ Alan Douglas 在"The Crystal detector"（晶体检测器）（*IEEE Spectrum*, April 1981）中告诉我们，Pickard 到了 1920 年，已经在实验中测试了 31 000 多种矿物质和导线的组合。

⑧ 1N26 在 24GHz 最大的转换损耗为 8.5dB。

早期的点接触二极管和现代的点接触二极管并没有多大的区别。但是因为封装技术进步了，特别是寄生电感得以减小，一个更主要的不同是半导体质量上的差别。单晶锗和单晶硅的晶锭直到20世纪50年代才有，所以早期的二极管都是用多晶晶片制造的。在多晶材料中载流子的输运不够好，所以电阻率对应于一个给定的掺杂浓度要高很多。多晶晶片被打碎成碎片，破裂的地方大多数发生在晶粒的间界。人们通过很多努力去寻找大晶粒内的材料，希望可以找到近似于单晶的样品。被选中的块放置在带螺栓的基座上，然后将晶片的边缘接地。在组装到一个封装中并调整触须线（通常是钨做的）之后，对二极管进行测试并根据质量好坏分类。这些二极管在性能上可能会有很大的差异，那些性能好的二极管被雷达工程师当宝贝保管起来。当然，现在的半导体器件的质量一致说来都很好，所以因成品率不高带来的损失和器件串联电阻都是非常小的。最近用GaAs做成的点接触二极管可以在170GHz的频率下工作，其钨触须线尖端的直径为100mm[9]。现代的深亚微米光刻技术可以制造非常小的接触，所以现在制作的大部分肖特基二极管都是利用了这个优势，在一个二极管的表面上形成一个微小接触点的阵列。然后将触须线用恰当的强度（为了有好的稳定性）压在其表面，以与阵列中的某个点形成接触。截止频率在10THz范围的点阵列器件已经试验成功，良好的性能主要是将多余的寄生材料都刻蚀掉了[10]。

因为肖特基二极管的工作不依赖于少子注入，就不会有反向恢复延时。而且，大多数现代的肖特基器件的理想因子 n 都非常接近于1，通常在1.05～1.1之间：

$$i_{\mathrm{D}} = I_{\mathrm{S}} \exp\left(\frac{v_{\mathrm{d}}}{nV_{\mathrm{T}}}\right) \tag{6}$$

在偏置为正向导通电流1mA时，室温下的动态电阻不会大于25Ω太多。图9.5给出了肖特基二极管的符号图。

图9.5 肖特基二极管符号

9.4 变容管

肖特基和结型二极管共同表现出的依赖电压的电容特性对于实现压控电路非常有用。因此，举例说，变容管在压控振荡器（VCO）中就有很广泛的应用。

我们已经看到对线性掺杂，电容随电压的改变而引起的变化不那么剧烈，而突变掺杂可以使得电容随电压的改变变化更加急剧。用后面这一种掺杂形式制作的二极管能提供的电容的典型调整率为2:1。虽然初看起来不可能取得比突变掺杂更为有效的电容随电压变化，采用超突变掺杂还可以改进电容对电压的敏感度。这种超突变掺杂是通过指数式掺杂分布来实现的。

对所有这些不同的掺杂分布，二极管电容对电压的依赖关系基本都还是遵循前面给出的公式，在这里再次列举出来：

$$C_{\mathrm{j}} = \frac{C_{\mathrm{j0}}}{(1 - V_{\mathrm{j}}/\psi_0)^m} \tag{7}$$

我们已经知道对于线性缓变掺杂和突变掺杂，m 分别为1/3和1/2。对于某些超突变

⑨ “Die MIM oder MIS Dioden”（http://mste.laser.physik.uni-muenchen.de/~mst/mim.html）。

⑩ http://info.iaee.tuwien.ac.at/gme/jb97/97_07.htm。

掺杂分布，m 可能超过 1(在商用的器件中，m 值超过 2 都存在)。所以，用超突变掺杂分布，电容值可以在一个量级范围内发生变化。这种二极管使得 LC 振荡器的频率调谐范围为 3∶1，因而比如说可以实现电调谐的调幅收音机。

超突变掺杂变容二极管的一个典型电压与电容关系如图 9.6 所示。

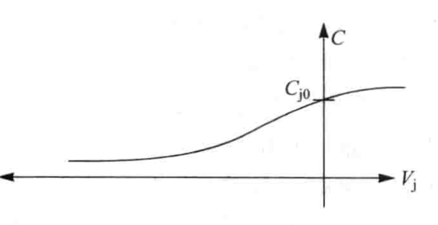


图 9.6　典型的变容管 V-C 曲线

参量放大

在“二战”快要结束的时候，MIT 辐射实验室发现了一个让人费解的现象：在某些情况下，一个普通的点接触二极管可能有能量增益，但又没有观察到有违犯能量守恒原理的情形，因为直流偏置和一个本地振荡器都可以作为能量源，但是人们还是不太清楚能量怎么从这两个能量源的其中一个或者两个转换到信号的频率上的。使用一个模拟计算机进行的后续研究发现，非线性的结电容是产生这一效应的主要原因。然而，除了这个关键性的观察之外，再没有其他的解释了[11]。

现在人们知道了，这种现象叫做参量放大。这个术语是从有意识地以一种设计好的方式改变电路的参数得到的。对于参量二极管，放大是如何发生的主要特征可以这样来定性解释：考虑一个简单的 LC 槽路，其中电容就是常见的平板电容，但是平板间间隔是可变的。在某个固定的平板间隔值下，提供一定量的初始电荷，这个槽路就在其自然频率上振荡起来。当然，如果损耗存在的话，那么振荡幅度会随着时间指数衰减。

现在考虑在电容的电压达到最大值时，将电容的极板距离拉大，会发生什么？在那一时刻，两个极板间的静电作用力达到最大，所以为了将极板分开，我们要对电场做功。那么我们做功的能量去了哪里呢？如果这个槽路没有损耗，那么这些能量都被存储在了 LC 槽路中，从而使得振幅增加。为了在不改变电路储存的能量情况下，将电容值恢复，我们在 90°相位之后，此时电容上的电压为 0，很快地将平板位置推回到复原的位置。再一个 90°相位后，又将极板分开，如此反复。在每个周期中用这种方法提供了两次能量，就可以不断地增加振荡的幅度，这就实现了放大。

注意我们改变极板距离的频率是 LC 网络振荡频率的两倍。前者叫做泵浦频率，在实际的实现中是由另一个振荡器提供。泵浦振荡器周期性地改变二极管的电容，使得在槽路的振荡频率上出现了负阻。

所以，如果我们没有违背能量守恒定律，那么参量放大器的能量从哪里而来？在标准放大器中，直流能量是最终注入信号能量的来源。在参量放大器中，能量从泵浦振荡器获得。就像我们在实验中必须要做功将电容的极板分开一样，这个泵浦振荡器也必须对参量放大器做功。因此，参量放大器就不属于线性时不变系统，所以可以将某个频率上的能量转化到另一频率上。

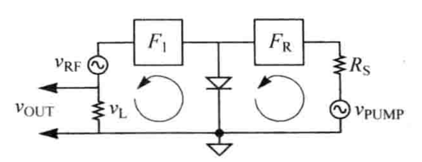


图 9.7　简化了的参量放大器

图 9.7 所示的简化了的参量放大器原理图更详细地说明了泵浦振荡器和主要的输入信号是怎样加到二极管上的，同时也给出了输出端口的位置。RF 输入 v_{RF} 通过带通滤波器 F_1 加到二极管上，这个滤波器调

[11] 在之前的 Torrey 与 Whitmer 的引文中(见脚注②)。

谐到输入信号的频率。同时，泵浦输入（其为输入信号频率的两倍）通过另一个带通滤波器 F_2 也加到二极管上，该滤波器调谐到泵浦信号的频率。泵浦信号调节二极管的电容大小，使其在 RF 输入频率上产生一个负阻。左边的回路就可以有一个净增益，进而在 R_L 上产生一个放大了的输出信号。

参量放大器一个有价值的特点是在放大的过程中只用到了电抗元件，因此可以得到非常低的噪系数声。事实上，在微波激射器（maser，即受激辐射的微波放大器，译者注）发明之前，可得到的最低噪声的放大器就是参量放大器。但是现在人们很少用到它了，因为常规的放大器也可以达到很好的工作性能，而且不需要倍频的泵浦信号。不管怎样，知道参数改变可以有放大作用是很有意义的。起码它告诉我们，利用两端器件也可以获得增益，这个结果也与下一节的内容相关。

本节中最后需要提到的是，分布式放大器（也称行波放大器）的构造是利用了参量二极管的。

9.5　隧道二极管

1957 年的夏天，年轻的 Leo Esaki 在成立不久的 Sony 公司进行自己博士论文的工作。他的题目是研究锗二极管的击穿电压随掺杂浓度不同的关系。因为“预期得到的结果”是广为人知的，这个论文题目起初看来并不是那么的有前途，拿一句老话来说做出来也大半是被科学界置于一旁。Esaki 自己也意识到了这种危险，所以他选择填补文献中的空白，他采用了极端的无人尝试过的高掺杂浓度，然后在一个很大的温度范围观察其特性与温度的关系。

首先，他证明了重掺杂会减小击穿电压。这并不新鲜。新奇的是：他观察到在非常重的掺杂条件下（简并掺杂），击穿电压可以被减小到小于 0。也就是说，在某些情况下需要利用一个不断增加的正偏电压来减小电流。在这个工作区域，这种重掺杂二极管呈现了负增量（或称微分）电阻的情况。当正偏电压进一步增加，这种二极管又恢复到通常的二极管的特点上。

如图 9.8 所示，正偏电压由 0 变为 V_P，隧穿二极管正向电流先达到峰值 I_P；随后电压由 V_P 变为 V_V，电流减少到谷值 I_V。在电压大于 V_V 的区域，二极管的性能又变得与普通二极管相同。除了简单地报道观察得到的结果外，Esaki 还做了其他的工作，他解释这种现象是因为量子力学中的隧穿引起的。[12] 在那个时候，隧穿还只是为了解释各种现象而提出的理论概念，没有人直接观察到过。Esaki 用如此简单的一个结构直接证实了隧穿，这个实验是如此地震惊，导致他获得了 1975 年的诺贝尔物理奖。

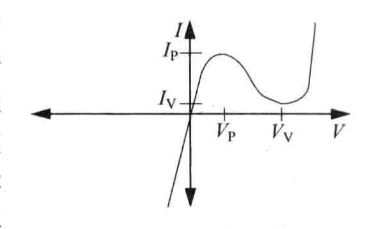


图 9.8　典型的隧道二极管特性

除了纯科学上的重要性，隧穿二极管具有的负阻效应也是其可能成为微波电路中提供增益的原件。假设一个由标准的电阻分压器构成的放大器，其中一个电阻为负值：

$$\frac{v_o}{v_{in}} = \frac{R_1}{R_1 + R_2} \tag{8}$$

[12] L. Esaki, “New Phenomenon in Narrow Germanium p-n Junctions,” Phys. Rev., v. 109, 1958, p. 603

其中 $R_2<0$，那么输出电压将大于输入电压，我们又构造出一个放大器。

商用的锗隧穿二极管，当正向偏置为 350mV 时，典型的谷电流值从几百微安到几毫安（有些可以高达 20mA，但那是非常罕见的）。峰值电流一般为谷底电流值的 5～10 倍，通常在正向偏置 65mV 附近取得。典型的峰值负阻电导范围从几微西门子到几百毫西门子。

一个重要的品质因素是在哪个频率范围内二极管可以产生净负阻。再次考虑我们基本的二极管模型（图 9.9）。外加调频原件，可以让并联电容 C_P 和串联电感 L_S 在任意频率都中和掉（起码在原理上是这样）。所以模型中只需要考虑的参数是 R_S，R_j（其值可能为负）和 C_j。平行的 R_jC_j 网络可以转化为串联的 RC 网络，其中电阻为

$$R=\frac{R_j}{1+(\omega R_jC_j)^2} \tag{9}$$

此时，净电阻为 R_S+R，在某个频率极限，净电阻的值将变成负数。这个电阻截止频率表示为

$$\omega_{\max}=\frac{1}{R_jC_j}\sqrt{1-\frac{R_j}{R_S}} \tag{10}$$

一个好的隧穿二极管使得结电容（和寄生电容）以及串联电阻达到最小值。

负阻的另一个明显的用途是在实现振荡器上。假设二极管的负阻超过其平行 RC 网络的正电阻，那么将一个隧道二极管与共振器相连就构成了振荡器，如图 9.10 所示。在这个简单的振荡器中，RLC 网络（电感直接与直流偏置相连）是一个有损的共振器。隧穿二极管与共振槽平行，并将它的偏置调节到负阻区域。只要槽的耗散可以被二极管的有效电阻抵消，那么振荡将持续产生。在很多高频应用中，共振槽都不是由集总的 RLC 构成，而是共振腔（包含谐振传输线的那一类）。

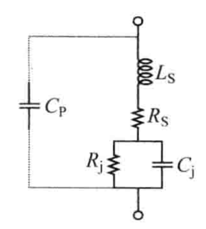


图 9.9　隧道二极管的简单高频模型

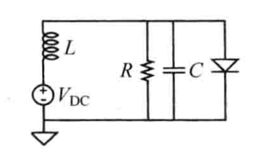


图 9.10　简化了的隧道二极管振荡器

因为用到的极端掺杂情况并不要求非常纯的材料或者在制造过程中非常的干净，因为隧穿二极管结构简单，很高的工作频率容易制造，很多人都乐观地认为，在大量的应用中可以代替晶体管。全世界的实验室都在尝试使用隧穿二极管构造微波系统甚至是计算机。

但是，读者可能意识到了隧穿二极管在现代的微波技术中并不占据主导地位。一个原因是隧穿二极管并不能像理论中的那样，作为一个两端器件，构造单方向的放大非常困难。因此，隧穿二极管电路一个非常不好的特点是：一旦负载的阻抗改变，其他的一切直到输入都将被改变。这个特点使得级联非常困难。所以，除非是非常简单的，或者独立的电路，隧穿二极管都是不适用的。另一个原因是负阻区域只跨越了一个非常小的电压区域，因此有用的电压值被限制在了很小的范围，导致典型的信号输出能量小于 1mW。最后，增益随着

偏置和温度的改变保持恒定是很重要的。所以,制造厂都放弃了生产,今天能够找到的少量的隧穿二极管都非常贵,这也反映了工业界不愿意制造它们。估算一下,1999 年全世界售出的隧穿二极管少于 10 000 个。不过,隧穿二极管仍然是进行学术研究的好课题,关于它的出版物也是持续不断。

反向二极管

反向二极管只是将隧穿二极管的阴极和阳极相互交换,而且采用的掺杂浓度使得电流的峰值和谷值基本相等。*V-I* 曲线如图 9.11 所示。将阴极和阳极相互交换后,二极管的正向压降变得很小,起码在正向特性上与理想二极管的特性非常接近。但是,注意到反向二极管的反向击穿电压小于 1V,这种二极管只能用作电力整流器。这种二极管在小信号检波器或者混频器中偶尔使用。

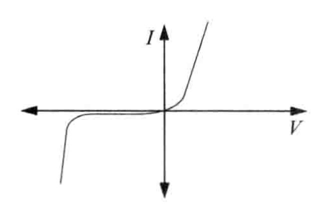


图 9.11 典型的反向二极管特性

9.6 PIN 二极管

PIN(p-本征-n)二极管尝试解决标准二极管遭遇的少子存储问题。(在这里是一个非常重要的题目:如果我们多考虑一些的话,某处遇到的问题可能在其他地方成为优势。)就如名字表示的那样,PIN 二极管是将一个掺杂浓度非常低(本征,或者近似本征)的半导体层插入到标准结的半导体层之间。因为中间插入层的电阻率高,它决定了整个二极管的电阻。当二极管正偏时,注入的电荷使得本征区域的电导率增加。

在低频下,PIN 二极管与普通二极管的特点相同,因为复合过程非常迅速,所有的注入电荷在终端缓变电压变化的周期内就会被消灭。但是,当掺杂很轻的时候,载流子的寿命特别长,所以二极管不能跟随外接电压的快速变化(大于等于 MHz 量级)而变化。所以,电荷浓度基本保持不变,使得 PIN 二极管像一个电阻。进一步的,这个阻值是外接直流偏置和通过二极管电流的函数。

因为 PIN 二极管像一个压控电阻,它在 RF 电路中被广泛的用作调节器,可调衰减器和开关。因为它可以用标准的微电子工艺制造,PIN 可以被做得很小,相应的寄生参数也很小。滤波的器件可以连接到电路上(或者从电路中移除)来控制波段转换。自动增益控制(AGC)是 PIN 二极管另一个非常优秀的应用。图 9.12 所示的电路介绍了 PIN 用作衰减器和开关的典型用法。

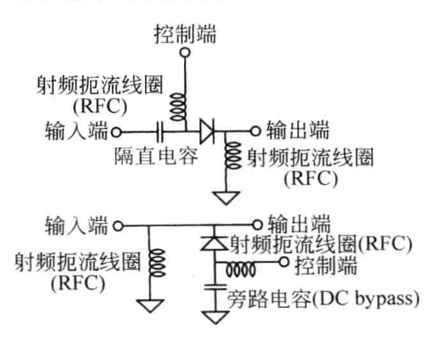


图 9.12 PIN 二极管衰减器/开关电路

图中的上面部分,PIN 二极管串联在输入和负载之间。二极管的偏置电流通过控制输入和在各种关注的频率阻抗都很大的 RF 扼流装置共同提供。隔直电容放置直流电流流入输入源。在输出端不需要隔离电容,因为接地端的 RF 扼流装置使得直流输出电压为 0。当偏置电流较低时,二极管阻抗高,几乎没有信号从输入传播到输出。偏置为 0 时,导通完全依靠二极管的并联电容。最大衰减的典型值为 20~40dB,导通发生的频率越高衰减将增加。当偏

置电流较高时，二极管的电导率很高，将输入和输出导通，带来的损耗小于1dB。在这个极限，PIN可以被当做开关。

第二个电路利用了并联的形式，使得PIN二极管为负载电阻添加了一个并联的电阻。所以，衰减随偏置电流的增大而增大。

这两个电路通常通过T形连接而连接在一起，中间有一个分路为任意的馈通原件提供通路。T形开关的隔离度很高，为70～80dB，但是其代价是最小的损耗也将增加。

应该提出的是，对基本的结构进行改变，可以得到更加复杂的开关装置，如单刀双掷开关(SPDT)等。这些都是留给读者的练习。

最后，因为PIN是最基本的压控电阻，所以它可以用在任意的低频情况作为分压计，例如压控滤波器和移相器(不止这些例子)。

9.7 噪声二极管

当一个足够大的反向电压加到半导体结上，电场将使电子从原子的束缚中脱离出来。如果电场和平均自由程足够大，那么这些自由电子就可能被迅速的加速，撞击原子，产生更多的自由电子。这些次生电子将再被加速，再撞击，而产生更多的电子，如此反复。一旦电压足够大，发生了这种雪崩击穿，电流就会异常迅速的增加。换一种说法，在击穿区域电压电流的相互关系不敏感。因为这个原因，这种二极管(通常被称为齐纳二极管，虽然齐纳击穿的机制与雪崩击穿机制并不相同)通常被用作基准电压或者稳压器。

早期发现表明齐纳二极管引入大量噪声(实际上这个噪声最先是在充满气体的真空管中发现的，此时的真空管同样利用了雪崩击穿而工作)。[13] 工程师需要通过大量的努力来保证这个噪声没有进入信号关键路径，而影响到电路的信噪比。

基于“生活给了你柠檬，那就做成柠檬水吧”这样随遇而安的态度，这个噪声源可用以提供宽带白噪声，这种白噪声的有效噪声温度为几千度，但是并不要求电阻真的达到这个温度。这种二极管因此被用作不可替代的安全、简洁的低功耗噪声源，一个典型的用处是用以噪声的测量。

这个噪声根本上是由雪崩效应的随机性引起的：电子移动随机的距离，随机地得到一个能量，撞击原子，随机地使得一些被束缚的电子成为自由电子。因此雪崩击穿的频谱非常宽。只要器件的寄生效应很小，那么输出噪声频率就会在一个很宽的频谱上。

一个问题是，不同二极管之间的噪声差异很大，如果它们被用作噪声测量的基准，那么每一个都必须通过统一的标准校正。另一些特点，例如转换时间、随温度性能变化等，都需要是可调节和可控的。人们又设计了亚表面和护环稳定齐纳二极管来减小随时间的不稳定度。由于一些不标准结构的影响和校正需要的大量人力，源自NIST(或者其他实验室)标准的噪声二极管非常贵(在写这本书的时候，价格为2000美元)。

图9.13 典型的噪声二极管

图9.13是噪声二极管的典型连接图。给二极管提

[13] 需要指出的是，充满气体的雪崩波导噪声源被用作实验室噪声标准，二极管的噪声源可以以此追溯。

供一个电流偏置，使其可以提供一个稳定的平带噪声谱(电阻 T 形网络起到阻抗匹配的作用)。RF 扼流高频时与地面连通，隔直电容(并不是在所有的设计中都有)防止二极管直流电压打乱负载接收到的噪声二极管的输出。

可以在基本电路上添加其他的原件使得电路更加精细，可提供白噪声频率范围更大。这些添加的电路一般是为各个二极管定制的。

9.8 急变(snap)二极管

我们继续使用柠檬水的理论，急变二极管(又叫做阶跃恢复二极管)是经过精心设计的 p-n 结，其主要目的是为了利用二极管的少子存储效应。我们回忆，给正偏的二极管加上反向的偏压时，并不是 p-n 结二极管突然关断，而是首先将注入的少子移除。在这之前，二极管还是开启的，导通电流引起一个很小的压降(例如 0.7V)。这个偏离理想情况的现象也表明，在射频电路甚至是开关式电源中，反向的导通使得效率明显降低。在一些极端的情况中，可能是某些元件失效。

急变二极管将这种不利的现象转化为了优点。它利用了存储的电荷被完全移除时会发生的由开启到关断的迅速跳变，因此可以利用急变二极管制造上升时间非常短的信号。根据时间和频率关系，上升时间越快也就意味着包含了更多的高频分量。急变二极管可以被用作谐波倍增器、抽样电路门控信号发生器和其他的快速电路中。Hewlett-Packard 是制造急变二极管的先驱，[14]他们现在还将急变二极管应用在几千兆赫兹采样示波器前端。利用急变二极管可以产生出上升时间为亚 10ps 的信号。

为了理解为什么急变二极管有用，我们考虑图 9.14 中的电路及其对应的波形。在这张图中，虚线为输入电压 V_S，实线为输出电压 V_{OUT}。如果二极管的反向恢复时间为无穷快，那么输出电压永远不会为负，在输入电压为负时，输出应为 0。但是，如果恢复时间不为 0，二极管就会持续一段时间导通，会有一个很大的反向电流直到存储的少子都移除为止。一旦这些电荷都被移除，二极管的电流会迅速变为 0，使得从开启到关断会有一个非常锐利的波形。在这张图中，反向恢复间隔正好占据了输入正弦信号的 1/4 个波形。

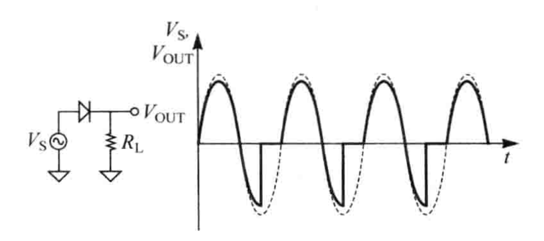


图 9.14　逆向恢复的图解

通过观察输出波形，从频谱中明显包含了谐波分量。对于一个周期性的输入，输出的频谱包含了输入信号的谐波分量，使得输出的频谱呈梳状，大部分的能量都集中在急变二极管

[14] A. F. Borr, J. Moll, and R. Shen, "A New High- Speed Effect in Solid- State Diodes," ISSCC Digest of Technical Paper, February 1960.

上升时间的倒数范围内。[15] 可以使用滤波器筛选出想要的频率分量，这样急变二极管可以当做倍频器使用。利用这种方法，100MHz 的输入信号可以产生出千兆赫兹的输出。另一方面，倍频器如果发生一些意外就很麻烦。一个常见的例子就是普通的电源整流器，一个 60Hz 的输入频率可能变化到高 AF 和低 RF 波段，造成干涉。这个现象在音频放大器和 AM 收音机中特别讨厌，此时干涉会导致声效失真。

另一种对于急变二极管作用的描述是：使得缓变波形变得锐利。在这种情况下，急变二极管可以用作高速取样器，一个缓慢的脉冲通过急变二极管电路加速，制造一个快速脉冲，驱动一个取样开关。

急变的速度取决于掺杂，因为少子被清除的速度取决于内建电场。重掺杂使内建电场大，因此急变的速度快，进一步使得电场最大的掺杂形式也是非常有用的。超急变变容二极管中采用指数变化掺杂就可以用来制造急变二极管。一般说来，变化指数 m 越大，二极管的变化速度越快。

既然击穿二极管也与掺杂有关，我们希望击穿电压低，关断迅速，其他参数都不变。这个关系在实际中真的被观察到了。最快的急变二极管也是低击穿电压的变容二极管。

反向恢复时间是正向电流的一个函数(存储的电荷越多，恢复时间越长)，所以频谱成分也是偏置的函数。梳型信号发生器和倍频器中，在图 9.14 的基本电路上经常加入调节偏置的装置。可以变化偏置和输入的幅度优化设计中的参数(例如，第十个谐波的数量)。将这些自由度和滤波器一起使用使得很多急变二极管倍频器提供的转化增益高达－15dB，甚至更高，而且直到达到带宽极限(可以通过带宽和上升时间的关系估算出来)，转变增益都基本保持不变。

最后指出关于急变二极管，不要相信关于反向恢复的 Spice 仿真。大多数制造商提供的二极管模型都没有抓住准确仿真需要的精细的效应。特别是电荷存储的方程从根本上可能就是错误的，提供的载流子渡越时间不完整。

9.9 耿氏(Gunn)二极管

思考振荡器如何工作的一种思路是将其看做是负阻和谐振腔的一个组合。低微波频率的负阻是通过晶体管连接一个正反馈结构(如 Colpitts 振荡器)来实现。但是在频率的极限时，有源器件造成的频移很大，使得振荡频率更多地依靠晶体管，而非谐振腔。在这个极限，振荡频率被限制在一个极限，f_{max}。

另一种方法是使用一个通过某些奇异的物理现象而形成负阻的两端器件。回忆固体中电子的有效质量跟自由空间电子的不同，就是固体材料中电子与其他电子相互作用的结果。其他的电荷，包括那些流动的电荷都是外加电场的结果。所以，有效质量可能随电场变化。

GaAs 的能带结构中，倒带包含了能量差异微小的能谷(InP 是另一个这样的材料)。这两个能谷中的电子有两个不同的有效质量，较重的电子在能量更高的能谷。在低场下，只有低能谷的电子参与导电。随着电场的强度增加，这个能谷被填满。直到电场的强度达到某个阈值(对于 GaAs，这个值在 300～350kV/m 左右)，此时上面的能谷也开始被填充。因为

[15] 旧的经验中带宽和上升时间的关系($f_{-3dB}t_r=0.35$)可以在此处近似使用。

电子质量更高，通过这个阶跃的偏移速度随着电场的增加而减小。因为电流密度与能量成正比，我们遇到了一个有趣的情况，只有某个范围内的电场强度，才能使电压和电流的增长关系相反。如图9.15所示，我们刚才描述的情况其实就是一个增加的负阻。这个现象最先被J. B. Gunn观察到，这个现象和器件也就一次命名。但是Gunn现象的理论预测只在前几年才被Ridley，Watkins[16]和Hilsum[17]等人真正实现。叫Gunn比叫他们所有的名字方便很多，所以人们还是习惯使用前者。

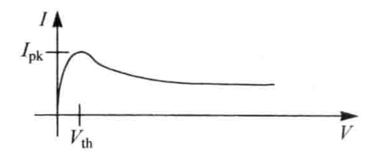


图9.15 耿氏二极管的典型的 V-I 图

Gunn二极管一个非常吸引人的特点是：它是一个体效应器件。人们叫它二极管基于一个普遍的常识，因为它是一个两端器件。Gunn二极管其实没有结。制作时不需要异常的过程也不需要昂贵的光刻。只需要简单地将其封装，将材料偏置到负阻区域，就可以使用了。

当在CW振荡器中使用时，偏置电流总是被调试到比产生最大能量输出的电流小10%～20%左右，与之对应的电压为阈值电压的3倍量级。在瞬时应用中，为了得到更大的瞬时输出能量，使用的电压比这个值还要大三倍。

商用Gunn振荡器的频率范围可达5～150GHz，[18]CW输出能量的范围也从几毫伏到几百毫伏甚至更多，DC到RF的转换效率约为10%。瞬时输出能量的值在任意频率都会高几百倍，Gunn可以承受的最大瞬时输出能量在千瓦量级。

对于任意的放大器和振荡器都必须考虑它们的噪声。一个常用的器件噪声参数叫做噪声量度 M。对于一个二端口器件，这个值表示的是器件的有效电阻与相同阻值的无源电阻噪声的比值：

$$M \overset{\text{def}}{=} \frac{\overline{v_n^2}}{4kT\,|\,R\,|\,\Delta f} \tag{11}$$

GaAs Gunn二极管在10GHz下测得的噪声量度在15dB。

在器件设计中，频率和能量上的性能总会做一个折中，这就使得输出能量与频率的平方的乘积会有一个范围。一个近似的但在直觉上让人喜欢的推导是首次观察到一些关键的器件尺寸局限了频率响应，所以，这些尺寸必须随着波长变化（或者与频率成反比）。但是，击穿电压（假设击穿电场为常数）也随着关键尺寸变化。能量与电压的平方成正比，所以与频率的平方成反比。在CW工作中，最大值可达500W·GHz^2，瞬时模式可达10^5W·GHz^2。这种与频率的平方成反比的关系，越靠近毫米波将恶化为一个立方反比关系。

偏置电压影响外加电压和产生的电流之间的相位延时，所以它将影响到Gunn二极管的有效电抗。这个电抗部分决定了产生谐振频率的谐振腔的结构。所以，这是非零的供给推进效应，如果需要将偏置电压设定到最小值的话，就应该将其稳定下来。换句话说，对于电源电压的敏感性为我们调节Gunn振荡器的频率提供了一种可能。

温度也会产生很强的影响，温度升高，阈值电压和输出能量电压峰值都会减小。典型的

---

[16] "Microwave Oscillation of Current in Ⅲ-Ⅴ Semiconductors," *Solid-State Commun.*, v. 1, 1963, p. 88.

[17] "The Possibility of Negative Resistance Effects in Semiconductors," *Proc. Phys. Sco. London*, v. 78, 1961, p. 293.

[18] "Transferred Electron Amplifiers and Oscillators," Proc. IRE, v. so, 1962, p. 185.

温度电压系数在－5000ppm/℃。[19] 如果温度变化，如何防止供给推进就会变得复杂。

9.10 金属-绝缘层-金属（MIM）二极管

另一种依靠隧穿的结构是金属-绝缘层-金属二极管。如果绝缘层很薄（在几纳米甚至更薄的量级，就像隧穿二极管一样，而且弄薄的原因也与它相同），就可能产生很大的隧穿电流。因为电压的非线性关系，MIM结构具有对称的非线性特性，如图9.16所示。在适当的偏置下（有可能为零偏），MIM二极管就像常规的二极管一样可以用于检波。因为电流通过绝缘是基于隧穿，而其他的电极又是导电性很好的金属，这种二极管可以工作在很高的频率。集成的MIM工作在光频都不是问题。

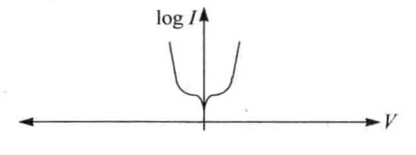


图9.16　对称的MIM二极管的具有代表性的V-I特性

9.11 IMPATT（碰撞电离雪崩渡越时间：IMPact ionization Avalanche Transit-Time）二极管

在1956年左右，Read提出雪崩击穿二极管可能产生负阻。他的简单理由是：使一个二极管工作在雪崩击穿状态，将其伸长到一定距离，就会产生一个不可忽略的载流子传输延时。因为这个延时，电流密度就可能落后于所加的正弦电流。如果这个之后可以达到足够长，就可以使电流在电压增加的时候减小。如果可以达到这样的效果，我们就可以构造出一个负阻了，它也就可以用在我们提到的所有可以使用负阻的应用中（隧穿二极管和Gunn二极管）。

受到Read的启发，几个研究小组就开始寻求将他的理论变为实际的方法。一些早期的工作表明非常容易实现。巧合的是，在20世纪50年代普通的制造工艺才能够生产出适合制造负阻的结构。就像Bell实验室的工程师在测试了满盒具有十年历史的二极管后所做的记录一样，他们只是"显得想要工作"。在那之后很短的时间内，频率在几十千兆赫兹的正当其可以通过非常简单的电路制造而成了（只包含二极管和调节二极管谐振频率的谐振电路）。

如图9.17中所示，IMPATT的雪崩击穿和漂移的区域是通过一个并联LC谐振电路调节的，这也反映了载流子渡越延时和工作频率之间的密切关系。在Read最初结构上做一些变化，可以在效率、调谐范围和工作频率等因素中做折中。但是，一般来说，IMPATT是一个窄带器件，典型的谐振频率只在测量值的百分之几十范围内。所以IMPATT必须要与安装的谐振腔匹配。粗糙的调谐可以通过调整腔的尺寸达到，精细的调谐（包括调制）可以依靠改变二极管的偏置电压实现。

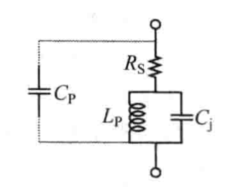


图9.17　IMPATT二极管的简单高频模型

20世纪60年代末期，可以工作在340GHz的IMPATT被生产出来。14GHz时，CW功率为5W；8.5GHz时的瞬时功率为30W；这些参数都使这种器件前景看好。峰值效率为10%～20%可与Gunn二极管的值相比。性能的极限比Gunn二极管好很多，CW silicon

[19] 1 ppm＝10^{-6}。

sources 限制在小于 3kW・GHz^2，瞬时值比这个值还要好三倍。

因为需要利用雪崩击穿，而且输出功率取决于输出电压的摆幅，所以 IMPATT 需要较高的工作电压，典型的值是 50～100V。为了取得更高的输出功率，也可以将 IMPATT 串联起来。

高频下高的输出功率看起来在很多应用上可以让 IMPATT 源代替传统的真空管源（如磁电管）。不幸的是，就像我们已经看见的那样，雪崩击穿是一个引入噪声的过程，所以 IMPATT 制成的振荡器有很大的相位噪声。这个确定就将它们局限在了很小一部分不要求噪声较低的应用上。IMPATT 典型的噪声量度超过 30dB，在某些情况下，特别是电压增加时，可能达到 50～60dB。一个共识是 IMPATT 振荡器的工作电压应比达到最大功率输出电压时的偏置小 15%，达到噪声量度和输出功率的一个折中。

就像我们在雪崩二极管中最后提及的那样，实验者们偶然发现了 IMPATT 的一个反常的工作模式，被命名为 TRAPATT(trapped plasma avalanche transit time，束缚等离子体雪崩渡越时间)。这个模式在超低频是有效的，可以达到非常高的效率。如 De Loach 提到的那样，瞬时输出功率在 1GHz 时为 400W，效率非常引人注意，为 60%。这个器件有与 IMPATT 相同的高噪声问题，TRAPATT 在工业上的应用很少。

9.12 总结

我们研究了很多种如何将两端器件设计为在 RF 和微波频率中发挥各种作用器件的方法。高频放大器、衰减器、调制器、开关、检波器、振荡器和短脉冲发生器都可以通过简单的电路得到。

9.13 附录 自制便士二极管和晶体收音机

虽然现在硅是主要使用的半导体，我们不能够忘记，还有其他很多有机的和无机的半导体材料。我们注意到 Braun 在 1874 年最先提出半导体的精髓就在于可以整流，这比半导体这个词被制造出来和收音机的发明都要早很多。他发现很多自然存在的矿物质，如 PbS，FeS_2，$Cu_3AsS_{3\text{-}4}$/$Cu_3SbS_{3\text{-}4}$ 和 $CuFeS_2$/$Cu_2S \cdot Fe_2S_3$ 等都不遵守欧姆定律，电压极性不同时导通电流不相等。固体量子物理对这个问题的解释在五十年之后，所以 Braun 不能为他实验中观察到的现象给出一个令人满意的理论解释。就是在今天，对于很多半导体混合物的定量解释还是含混不清的，主要是制造出足够纯净和完美的样品很困难。

30 年后，无线电时代即将到来，更为简单的第四族半导体因为偶然的机会初上舞台。在世纪之交时，Acheson 在合成钻石的时候不经意地合成了 SiC。SiC 虽然不及钻石珍贵，但硬度相同，因此他发现 SiC 的直接且长久的应用就是作研磨剂。Gen. Henry Harrison Chase Dunwoody 利用这种半导体的性能制成了无线电检波器。

几乎同时 Greenleaf Whittier Pickard 发现硅也是非常优秀的半导体，同时为硅无线电检波器申请了专利。这个成就是这种基本的半导体第一次被承认和使用。Pickard 在这个领域用尽了一切努力，测试了 30 000 种矿石和接触线的组合（奇怪的被叫做触须线），为他们中的很多申请了专利，如图 9.18 所示。由 ZnO 和 Cu_5FeS_4 组成的接触应用最为广泛。

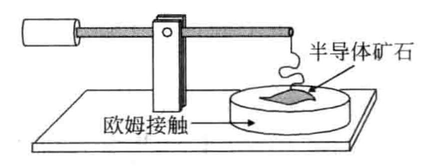


图 9.18 典型的晶体检波器

Pickard 把它叫做 Perikon 检波器，表示完美的 Pickard 接触(PERfect pIcKard cONtact)。Pickard 做科学家比当产品商人要当得好。

所有这些基于矿物的材料都用来制作晶体收音机。值得注意的是，这些晶体收音机通常工作时不需要任何超过入射无线电波的能量源。在最简单的形式中，晶体收音机包括一个天线、二极管和收话机，如图 9.19 所示。几乎所有适用的收话机都不需要使用可选择的电容，但是为了完整性和与那个时代的电路图的兼容性，我们还是将其包含在图中。

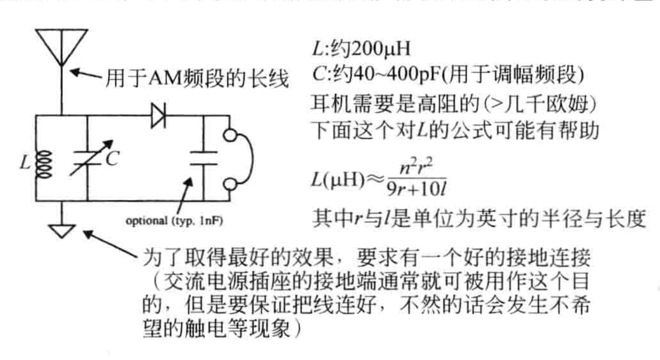


图 9.19　简单的晶体管收音机(未经优化)

为了弄懂一个简单的收音机怎么工作，我们必须要知道：入射波是调幅的，也即是说语音或者音乐都是对在空中传播的无线电载波的幅度进行解码得到的。调幅波的平均值为 0，我们的耳朵反应太慢了，所以不能对除了平均值之外的任何东西做出反应。二极管对这个对称的 AM 信号进行整流，制造出一个时域上的平均值与幅度成正比的非对称信号。

人类听觉系统不需要使用电池的原因是令人惊讶的。听觉的阈值是 1kHz，耳膜的移动距离与氢原子的直径相对应，所以，奇迹并不在收音机上，而是在生物学上。当然，缺少任何的电学上的放大也就意味着晶体收音机要求更好的天线和合理距离的广播站。

这些收音机中的标准晶体包含一块方铅矿(最好是含银的，虽然含有银，但是还是被叫做方铅矿是因为看起来像)，与这块矿的接触历史上是通过浸没在低沸点的合金——Wood 金中。Wood 金是铅、镉和铋组成的合金。这个接触一点也不关键，其目的仅仅是想与晶体形成一个低阻接触。Wood 金并不是必须的，只是简单的箝位就足够了。弹簧夹的钳夹就可以了。

另一个接触是非常关键的，因为触须线和晶体表面形成的接触是整流发生的地方。如果接触的压力过大，就会造成短路；如果太轻，电阻就会很大。触须线包含了非常细的线(例直径之后两毫米)，所以可以与方铅矿的表面形成最轻的接触。我们还必须在整个表面找到最合适敏感的接触点。因为接触点非常纤细，很容易被震动，所以需要经常调整。现在更容易做成一个好的晶体收音机就是因为能够生产非常优秀的不需要调整的二极管了。1N34A 型锗二极管(生产的厂家越来越少了)就特别适合做晶体收音机，因为它是一个封装好的单元，但是它就失去了触须线乱动的魅力。

方铅矿的替代品在二战中经常被士兵们使用——剃须刀的刀片。生锈的也可以工作，因为锈的主要成分是 Fe_2O_3，也是一种半导体(Pickard 在 1906 年发现的)。更好的还是美国安全剃须刀公司生产的“蓝色伙伴”刀片，因为它们包含了一层高质量的薄氧化层。用削尖的铅笔芯或者安全针做成的简易触须线就用在“散兵坑”收音机中。

人们发现我们只使用便士就可以制造出一个很好的二极管。实际上,这种二极管比剃须刀刀片做成的二极管要好很多,其性能基本与商店买的二极管相同。

在20世纪20年代早期,Grondahl和Geiger(不是那个Geiger)偶然发现Cu_2O也是一种半导体(而且它总是p型的)。将这个发现与其他早期的关于半导体的工作分开的原因是:铜氧器件在表面具有统一的性能,使得制造大面积的接触(而非点接触)成为可能。虽然在高频应用中并不希望大面积的接触,但是这是高电流的应用所需要的。因此,这种材料首先用在电源整流器中,代替了可靠性差的、体积过大的和能耗很大的真空管。但是,制造过程中需要约1010℃(1850℉)的高温,使得这种氧化物不能在家里制造。进一步,我们粗略看看Cu/O系统的相图就可以发现,气压和温度都必须控制在一个很窄的极限(温度窗在20~30℃),才能保证得到的是Cu_2O。如果偏离了温度窗,就会形成CuO。那时候的文献(多数是出自Walter Brattain之手,他后来与他的同事Joseph Becker一起因为共同发明了晶体管而获得诺贝尔物理学奖)反复强调,CuO是杂质,而且讨论了很多种避免他的方法(例如喷砂、碾碎和电解法)。

但是,作者做的小实验表明,CuO可能是杂质,但是这种铜的另一种化合物确实也是半导体(事实上,经常是n型的),我们家里就可以很简单的合成。在大气压下,只要温度低于1000℃,就可以合成CuO。你不需要超净间,在脏的房间也可以合成。

制造便士二极管的方法如下:

(1) 找到一个1983年后产的便士。因为在那之前制造的便士是用的均匀的铜锌合金(比率95/5),很显然的,锌将弱化我们的操作。新的便士基本都是用锌做的,只是在表面电镀了一层厚度为10~15μm的纯铜。(这个变化其实是发生在1982年中期,但是这两种便士在外观上并不能区别,为了保险起见还是选择1983年及其以后生产的。)

(2) 将便士彻底的清洁,但是轻轻使用铜罐清洗剂,并用旧牙刷将其擦亮,散发着铜特有的光芒。不要用研磨剂(记住,表面的铜非常薄)。只拿住便士的边缘也是可以的,这样就可以避免皮肤上的油渍污染便士的表面。最好是戴手套。或者不戴?谁知道呢?有可能汗液是关键的掺杂剂。

(3) 清洗和晾干之后(可以用手纸擦拭,加速晾干),将便士放在甜酥饼干上(折起来的铝箔也可以),放进预先加热到至少500℉(使用你的烤箱允许的最大值)的烤箱,烘烤15~30min。

(4) 关闭烤箱,并经其冷却。检查这个便士是否已经被一层均匀的深褐色的薄膜覆盖。顺便说一句,这个便士仍然是合法货币。你只是加速了它的自然氧化。

你可能希望一次同时烘烤很多的便士,而不是一个。它们很便宜(用超级计算机进行大量的计算表明每个只花费一美分:D),最好是让烤箱的能量发挥最大的作用,而且这样增加你找到活样本的机会。

找到一个触须线,一根细的磷青铜线就是很正规的选择。但是,我们也可以使用弯曲的安全针,虽然我们要极小心地使用最长的接触。铅笔芯更好,because it has a little more "give" to it,使得结构对接触的压力不敏感。一个0.5mm HB的铅笔芯就可以很好的工作,其他的也可以。使用这样的电极时,形成接触的时候需要与表面有一个角度,而不只直接的迎面相接,所以电极的边缘与便士的表面相接。也可以使用灯泡中得到的钨丝。

为了形成欧姆接触,夹着这个便士放在一块一端或两端被木螺丝钉固定的小木片上。

紧紧固定的螺丝钉提供的压力比让便士形成良好接触的压力更大。在螺丝钉上连接一跟导线。既然用前面提到的方法产生的 CuO 都是 n 型的，那么螺丝钉连接的一端是阴极，触须线连接的一段是阳极。

从图 9.20 可以看出，一个安全针和几转硬线可以将一小段铅笔芯固定。如果使用两个垫圈让安全针可以随意的转动(如图中所示)，那么就有足够的自由度来调试铅笔芯接触的位置和压力了。就像早先建议的一样，我们可以不用铅笔芯，用安全针作为触须线。但是，此时必须要非常注意只使用最轻的压力。即使是那样的注意，你可能还是会发现调整是需要技巧的，而且很不稳定。

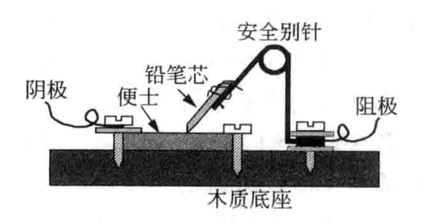


图 9.20　便士硬币二极管(侧视图)

现在，我们可以将这个二极管应用到晶体收音机中了。开始使用一个 1N34A 锗二极管对广播站信号进行收听，因为同时处理太多的变量可能很困难甚至造成失败。当广播站的信号被收听到，就将锗二极管断开，并用便士二极管替代它，其他的所有设置都不变。现在开始比较有意思的部分了，在表面寻找一个好的点，调动触须线的压力达到最大的敏感度。然后，即使所有都正常工作，最轻的移动也可能让一切停止工作(或者就是没有任何原因的自动停止了)，强迫你再从头做一次。这种性能可能是迷人的，令人愉快的，也可能让人抓狂和分心，这取决于你的性格。如果是后一种情况，那么就一直用 1N34A。

另一方面，如果你属于前者，一旦你体会到了便士二极管的所有乐趣，你可能很有兴趣尝试使用其他的半导体。去你本地的矿产点，使用表 9.1 列举的部分名单(和在 Braun 研究中提到的)作为指导，帮助你从头开始制造二极管。

表中的半导体，方铅矿是迄今为止最好的。它非常容易得到，相对来说也不贵(我最近买了一个高尔夫球大小的，搜寻到大概 5 美元，不过你的里程可能改变)，也都存在最敏感的点。一般来说，如果想取得最好的结果，需要使用非常细的导线制成的触须线。最好的触须线趋向于磷青铜线，并不是因为它们的化学成分，仅仅是因为它们弹性刚刚好，适于得到合适的接触压。但是，从废旧的发电机(比如微处理器风扇)得到的细导线，只要保持清洁，就可以良好的工作了。

如果希望找到其他的材料，需要注意到早期的研究人员在他们的寻找中，都不太清楚不完美的接触是很重要的，所以，寻找金属型的化合物材料，你从其中发现整流特性是极有可能的。好的选择就是不同金属的氧化物和硫化物(你可以通过测试表 9.1 来证明)，或者尝试着合成它们。找到一个钉子，让它生锈。拿一大块长期暴露带有铜绿的铜。如果你的家人不同意你用不满铜绿的古董做实验，那就将一些便士扔到盐水里，就可以得到自己的铜氯化物了。

表 9.1　一些天然的半导体材料

名　　称	化　学　式	极　　性
方铅矿	PdS	可变
黄铁矿	Fe_2S	
氧化铁	Fe_2O_3	p 型

续表

名　　称	化　学　式	极　　性
金刚砂	SiC	可变
黄铜矿	CuFeS	
氧化亚铜	Cu_2O	p 型
氧化铜	CuO	n 型
黄铜矿	Cu_2S	
红锌矿	ZnO	n 型
硬锰矿	MnO_2	

试着烤烤其他国家带铜的硬币。实际上，我曾经得到的最好结果是韩国的十文硬币。质量和活性区域都远好于便士。5 文的硬币也是同样的材质，所以我猜它也可以很好的工作(在写书的时候，5 文硬币的价值少于一美分，这也使得它是最便宜的材料)。我试过的其他硬币包括新台币、德国的 1%马克、新加坡硬币和丹麦的硬币。它们都没有便士的工作性能好，但是可以工作的区域还是可以找到的。

最后，在这一部中，必须使用一个特殊的收话机或者耳机。普通的耳机，如个人立体声，都是用于调幅，而且阻抗为 32Ω，所以这里需要更高阻抗的耳机。幸运的是，这些(和锗二极管)都可以从以下渠道便宜的买到：

Antique Electronic Supply

6221 S. Maple Ave.

Tempe, Arizona 85283

电话：480-820-5411

传真：800-760-6789(美国，加拿大)480-820-4643(其他地区)

网址：http://www.tubesandmore.com

邮箱：info@tubesandmore.com

他们有小压电耳机，仅售 2.5 美元(编号 P-A480)和 15.25 美元的听筒(部分号码 P-A466)。这两种耳机都可以与本文中描述的晶体收音机电路兼容。耳机和听筒都有足够的内部电容，并不需要其他的并联电容。

第10章 混频器

10.1 引言

有关电子工程的教学中,对线性时不变(linear time-invariant,LTI)系统的研究投入如此多的关注,以至于使人们很容易得出再没有其他类型系统的结论。即使是承认有其他类型系统存在的话,违反 LTI 假设的情形也通常被认为是所不希望的。比如,小信号分析就是为了避免非线性带来的麻烦。然而高性能的现代通信设备实际上至少取决于一个不满足线性时不变要求的部件:混频器。超外差式①接收机采用混频器来完成信号的一个重要频率变换。Armstrong 的发明在 75 年期间一直是占支配地位的系统结构,因为频率变换同时解决了许多问题②。

在图 10.1 所示的结构中,混频器把一个接收到的 RF 信号变换成一个较低的频率③,称为中频(IF)。尽管 Armstrong 最初寻找这一频率降低的目的只是为了更容易地获得所要求的增益,但它也有其他许多显著的优点。比如,现在调谐是用改变本地振荡器的频率而不是改变多极点带通滤波器的中心频率来实现的,因此我们不是去前后同时调节几个 LC 网络以调谐到所希望的信号,而只是改变一个 LC 组合以改变本地振荡器(LO)的频率。因此 IF 电路就可以采用固定的带通滤波器④。作为一个结果,选择性是由这些固定频率的 IF 滤波器来决定的,实现起来要比可变频率的滤波器容易得多。此外,系统的总增益分配在许多不同的频带(RF、IF 和基带)上,所以可以得到所要求的接收机增益(典型的功率增益为 120~140dB)而不必过多地担心寄生反馈回路中可能产生的振荡。这些重要的特性解释了为什么超外差结构在其发明几乎一个世纪之后仍然占支配地位。

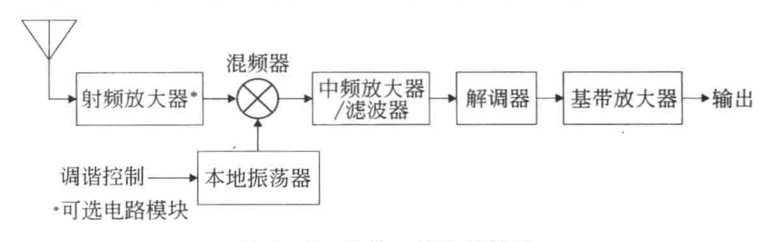


图 10.1 超外差接收机框图

---

① 为什么是"超"外差?原因是 Reginald Fessenden 已经有了被称作"外差"的某个发明,所以 Armstrong 不得不换个名字,以示区别。

② 这再一次证明了这么一个说法:成功有许多父亲(而失败则是一个孤儿,以便完成 John F. Kennedy-肯尼迪-的名言),Lucien Lévy 和 Walter Schottky,以及其他一些人,宣称是他们提出了超外差的概念。

Armstrong 的确不是提出超外差理论的第一人,但他是第一个认识到这个方法能干净利落地解决好多棘手问题的人,也当然是第一个奋力推进超外差理论实现的人。Schottky 最终放弃了他这个历史性的宣称,但 Lévy 却将愤愤不平带进了坟墓(然后几乎被遗忘了)。

③ 实际上,也可以采用将频率变高。

④ 认识到可以采用固定的中频使 Armstrong 的版本区别于 Lévy 的版本。

10.2　混频器的基本原理

由于线性时不变系统不可能输出在输入信号中没有的频谱成分，所以混频器必须或者是非线性的或者是时变的元件，以便进行频率变换。从历史上看，已使用过基于各种原理工作的器件（如电解质电池、磁性带、生锈的剪刀、真空管、晶体管和脑组织），这表明事实上任何非线性元件都可以用作混频器[⑤]。

当前采用的所有混频器的核心部分是将两个信号在时域中相乘。乘法的基本作用可以通过观察以下三角恒等式来理解：

$$(A\cos\omega_1 t)(B\cos\omega_2 t)=\frac{AB}{2}[\cos(\omega_1-\omega_2)t+\cos(\omega_1+\omega_2)t] \tag{1}$$

因此，两个输入信号相乘会在和频和差频上产生相同功率的输出信号，它们的幅值正比于 RF 和 LO 信号幅值的乘积。因此，如果 LO 幅值不变（通常就是如此），那么在 RF 信号上的任何幅值调制都会传递给 IF 信号。根据类似的机制，调制从一个信号到另一个信号之间的不被希望的传输也会通过在混频器和放大器中非线性的相互作用而发生。这种情况称为交叉调制，而通过改善线性度来抑制交叉调制则是电路设计中的一个重要考虑。

在已经认识到相乘在混频器中的基本作用之后，我们现在就列举并定义混频器最重要的特性。

10.2.1　转换增益

混频器的一个重要特性是转换增益（或损耗）。它定义为所希望的 IF 输出与 RF 输入值的比。因此对于公式(1)所描述的乘法器，变换增益就是 IF 输出即 $AB/2$ 除以 A（假设它是 RF 输入的幅值）。因此在这一例子中转换增益为 $B/2$，即 LO 幅值的一半。

转换增益如果表示成功率比，那么在有源混频器中它可能大于 1；对于无源混频器一般最好情况下也只是有电压或电流的增益[⑥]。超过 1 的转换增益通常是很方便的，因为混频器能在频率转换的同时提供放大作用。然而这并不一定能使灵敏度改善，因为噪声系数也必须加以考虑。由于这个理由，尽管无源混频器存在着转换损失，但它们在某些情形中可以提供极好的性能。

10.2.2　噪声系数：单边带(SSB)与双边带(DSB)

噪声系数的定义正如我们所预想的那样：它是输入(RF)端的信噪比(SNR)除以输出(IF)端的 SNR。然而这里有一个重要的细节常常使缺乏经验的，以及相当一部分从事实际工作的工程师犯错误。为了理解这一困难，我们首先需要进行一项重要的观察：在典型混频器中，实际上有两个输入频率会产生一个给定的中间频率。一个是所希望的 RF 信号，而

⑤　当然，某些非线性元件会比另一些工作得更好些，我们会集中讨论那些更为实际的种类。

⑥　有时作为例外引用的是参量变频器或放大器，在这类器件中，LO 的功率通过电抗非线性相互作用（一般用变容管）传输到 IF，从而使功率增益成为可能。这个事实告诉我们，这种系统更适合于归类为有源混频器。

另一个称为镜像信号。就混频器而言，这两个信号常常合在一起称为边带。

存在这样两个频率的原因是 IF 只是 RF 和 LO 频率之差的绝对值。因此，无论是高于或者低于 ω_{LO} 一个 IF 频率的信号都将产生同样频率的 IF 输出。因此这两个输入频率相隔 $2\omega_{IF}$。举一个具体的数值例子：假设系统的 IF 为 100MHz，而我们希望通过选择 1GHz 频率的 LO 来调谐到 900MHz 的信号。除了所希望的 900MHz 的 RF 输入以外，在 1.1GHz 的镜像信号也将在 100MHz 的 IF 处产生一个差频分量。

镜像频率的存在使噪声系数的计算变得复杂，这是因为来源于所希望的频率和镜像频率的噪声都变成了 IF 噪声，尽管在镜像频率中一般没有任何所希望的信号。在通常情形下，所希望的信号只存在于唯一的一个频率上，此时我们所测量的噪声系数称为单边带噪声系数(SSB NF)。在相当罕见的情形下，无论是“主”RF 还是镜像信号都包含有用的信息，这样就有了双边带(DSB)噪声系数。很显然，SSB 噪声系数将大于 DSB 的情形，因为两者具有相同的 IF 噪声，但前者只在一个边带中有信号功率。因此，SSB NF 一般将比 DSB NF 高 3dB[⑦]。遗憾的是，尽管 DSB NF 几乎不能作为任何通信系统的合适的品质指标，但有关 DSB NF 的报道却很多，其原因就是由于它的数值较小，因此错误地造成了它有较好性能的印象[⑧]。常常在谈及一个噪声系数时并不指出它是 DSB 还是 SSB 值，在这些情形中我们通常可以假设所引用的是 DSB 的值。

混频器的噪声系数往往远高于放大器的噪声系数，这是因为来自所希望的 RF 以外频率的噪声也可以混频到 IF 中去。SSB 噪声系数的典型值范围是从 10～15dB 或更高。正因为主要是混频器的噪声较大，我们才在接收机中采用了 LNA。如果 LNA 具有足够的增益，那么信号就将被放大到充分高于混频器及后面各级的噪声电平，所以整个接收机的 NF 将由 LNA 而不是混频器来决定。如果实际中混频器的噪声并没有那么大，那么对于 LNA 的需求也就明显减少。

10.2.3 线性度和隔离度

在现代高性能的通信系统中，动态范围的要求是非常严格的，在许多情形中常常超过 80dB 并且接近 100dB。正如在前一章中所讨论的，下限是由噪声系数确定的，它提供了有关多小的信号可以被处理的信息，而上限则是由大输入信号引起的一些严重的非线性起始点来确定的。

与放大器一样，压缩点是这一动态范围上限的一个度量，并且也以相同的方式定义。理想地，我们希望 IF 输出正比于 RF 输入信号的幅值，这是我们在有关混频器讨论中对“线性度”含义的解释。然而正如与放大器(以及事实上任何其他实际系统)一样，现实的混频器存在某一限制，超出了这一限制，输出与输入之间的关系就会是亚线性。压缩点是发生偏离理想线性曲线达到一个定量的量度时的 RF 信号值[⑨]。通常规定为 1dB(或者，少数情况下

---

⑦ 3dB 的差别是假设到两个边带的等转换增益相同。尽管这个假设通常能够满足，但是不一定总是这样的。

⑧ 两个重要的在双边带都包含有用信息的情形是射电天文(就像在宇宙背景辐射的测量中——大爆炸的回波)和直接变频接收器。我们在第 14 章会重新对 DSB 相对于 SSB 的解释进行讨论。

⑨ 一些制造商(以及一些作者)给出一个“输出压缩点”。如果在这点上的转换增益是已知的，这个数值可以被反演至输入点。遗憾的是许多人坚持把这个信息放在不易找到的地方，从而很难对混频性能作出客观的比较。我们将总是会明确地说出这个值是一个输入还是输出参数。

为 3dB)。我们应该标明这一压缩发生时的输入或输出信号的强度,以及转换增益,来允许对不同的混频器进行客观的比较。

双频正弦信号的三阶交调点(third-order intercept)也用来表征混频器的线性度。双频正弦信号的相互调制是评价混频器性能的一个关联的方法,这是因为它模仿了实际的情形,即一个所希望的信号与一个可能的干扰(也许就在一个信道频率之外的频率上)同时加入到混频器的输入。理想情况下,两个相叠加的 RF 输入中的每一个都进行频率变换而不会相互干扰,然而实际的混频器将总是显示出某些相互调制的效应,因此混频器的输出将包含三阶交调分量经过频率变换后的部分,这些未经混频前的频率为 $2\omega_{RF1}\pm\omega_{RF2}$ 及 $2\omega_{RF2}\pm\omega_{RF1}$。差频项可能会被外差成为 IF 通带内的分量,因此它们一般是比较棘手的分量,而和频信号则通常可以被滤波掉。

作为偏离线性混频特性程度的度量,我们可以画出所希望得到的输出和三阶 IM 输出与输入 RF 功率大小的关系,三阶交调点是这两条曲线外推线的交点。一般来说,交调点越大,则混频器的线性度越好,同样我们应当说明这一交调点是折算到输入端的还是直接是输出的量,以及转换增益,以便在混频器之间进行客观的比较。此外,我们习惯上把三阶交调点简写成 IP3,或写成 IIP3 或 OIP3(分别指输入三阶交调点和输出三阶交调点)。图 10.2 对这些定义作了总结。

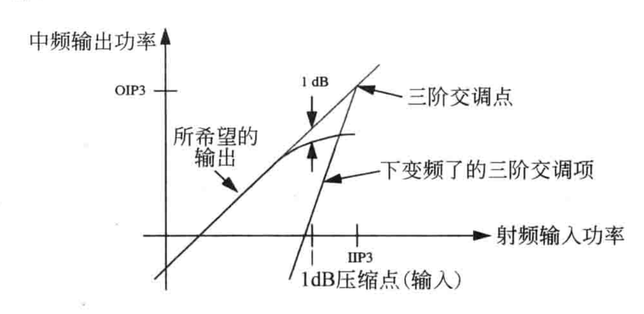


图 10.2　混频器线性度参数的定义

三次非线性失真也会对单个 RF 输入产生问题。作为一个具体例子,考虑制作一个低成本的 AM 收音机。不巧的是,AM 收音机 IF 标准是 455kHz(主要由于历史原因造成)。如果调谐到 910kHz 的电台(一个合法的 AM 无线电信号频率)要求 LO 设为 1365kHz[⑩]。三次非线性可能产生一个频率为 $2\omega_{RF}-\omega_{LO}$ 的分量,它在这一情形下恰好与我们 455kHz 的 IF 一样。

我们也许想去断言这样的分量不是一个问题,因为它可增加到所希望的输出中。为此我们甚至想考虑去利用它。然而,三阶交调乘积项的幅值并不与输入信号幅值成正比,因此它们可能会引起幅值的失真,从而破坏“正确”的输出(毕竟我们这里谈及的是幅度调制信号)。

即使没有发生前面例子中数字碰巧是产生了相同的中频项,各种不同的三阶交调项的频率还是有可能处在 IF 放大器的通带之内,这最终将使信噪比或信号失真比降低。

---

⑩　本振频率为 455kHz 也是可以的,但这是一个实现起来比较差的选择,因为这样一个在“低端侧的频率注入”会要求本振的调谐范围比起 LO 频率在所需 RF 信号频率之上的本征调谐范围更大。

另一个具有重要实际意义的参数是隔离度。一般希望在RF、IF和LO端口间的相互作用减到最小。例如，由于与RF信号功率相比，LO信号功率一般都非常大，因此送入到IF输出中的任何LO都可以在信号处理链的后续各级中引起问题。如果IF和LO频率很接近，那么滤波器的效果就不会那么有效，于是这一问题就更加严重，在许多情形中甚至反向隔离也是很重要的，因为反向隔离很差就会使很强的LO信号（或者它的谐波）回馈到天线中，从而可能被发射出去并干扰其他接收机。

10.2.4 杂散信号

就混频器本身的功能而言，它甚至可以混合出我们从未想要混合的各种频率分量。例如，某一（希望有的或不希望有的）信号的谐波有可能处在（或产生在）混频器系统的通带上，然后与本振信号（以及它的谐波）发生拍击（即混频，译者注），所产生的某些分量还可能最终出现在IF通带上。这些最终从IF系统输出的信号称为杂散响应，或简称为杂散(spur)。估计混频器的杂散信号在原理上十分直截了当，但做起来非常繁琐。现在已有了完成这个任务的软件工具，可以免去这一冗长乏味的工作，但描述一下这个过程仍然是很有意义的[11]。

设 m 和 n 分别为RF输入与LO频率的谐波次数，因此出现在混频器输出端（在任何滤波器之前）的杂散信号为

$$f_{\text{spur}} = mf_{\text{RF}} + nf_{\text{LO}} \tag{2}$$

这个公式看似简单，然而却有些误导：因为它必须对 m 和 n 的各种组合和符号都计算一遍，其范围一直到你需要考虑的最高次谐波。使这一步骤变得更麻烦的是，我们实际上必须考虑比额定输入通带更低的信号——至少要低到通带低边沿频率除以最大的 m 值。我们还必须考虑高于通带上边沿的输入频率。由于任何输入滤波器都不是完全理想的，因此LO的谐波仍然有可能与高于输入频带的RF信号相混频而在混频器的输出端产生带内杂散。

因此针对每一组 (m,n)，检查一下杂散信号的频率，然后决定它是否处在IF的通带范围内或者离它足够近，以确定是否值得做进一步的考虑。对每一个这样的杂散信号，需要返回去找到相应的RF输入频率，并估计在该频率处是否会因有一个足够强的信号而引起麻烦的可能性。然后根据需要对输入滤波器作恰当的修改以避免出现这些问题。

进行这一过程时有时假设为最坏情况：即在RF的输入端没有任何滤波。在这种情形下计算量很快就会变得非常大。但只要有足够的耐心，所产生的信息可以用来指导输入滤波器的设计。

举一个具体例子。假设我们希望设计一个FM接收器的混频器，它的额定输入通带包含88.1～108.1MHz范围的信号。对于10.7MHz的IF，要求LO能够从77.4MHz调谐到97.4MHz(假设低端注入，即LO比RF要低)。为了使说明的数字简单一些，我们有点不切实际地假设IF系统具有约200kHz的额定带宽。进一步假设我们不必考虑高于三次

---

[11] *AppCAD* 是一种可进行这种运算（以及其他一些对RF和微波工程师很有价值的功能）的优秀（并且免费）的软件。该软件源自惠普(HP)，现在属于安捷伦(Agilent)。有DOS版本和更新的Windows版本。软件 *RFSim99* 中也有杂散搜索工具。

的谐波[12]。根据这些假设，我们可以列出表 10.1。

表 10.1　FM 无线电例子中的杂散信号

m	低 f_{RF}	高 f_{RF}	n
−3	73.8	93.9	3
−2	72.0	92.1	2
1	88.0	108.2	−1
2	82.7	102.8	−2
3	80.9	101.0	−3

从表 10.1 的第一行，我们可以看到，在 73.8～93.9MHz 频带上 RF 信号的三次谐波可以与 LO 的三次谐波混频以产生在 10.6～10.8MHz IF 通带上的信号。注意在这个情况下输入滤波最多只是部分有效，这是因为大部分的杂散输入频段都与所期望的 FM 射频频段相重叠。如果在这个杂散的频带内确实有很强的干扰，我们唯一的选择就是改善 LO 频谱的纯度以减小它的三次谐波。这一改善对表中最后一行所示的杂散信号问题也有好处。同样，减少 LO 中的二次谐波则是避免杂散信号表中第二行和第四行所示问题的唯一实际可行的办法。第三行并不是不希望有的，它是描述了期望的接收机工作模式，之所以列在表中只是为了完整起见。

通过对所关注的系统完成了这种费力的分析过程，我们就有可能判断出系统对各种非理想情况的敏感程度，从而估计出是否有必要进行修正。

10.3　非线性、时变与混频

我们现在考虑如何实现作为混频作用核心的乘法功能。某些混频器可直接实现乘法，而另一些则通过非线性间接地实现乘法。我们沿着一条历史路径并首先考察一般二端口网络混频器的非线性[13]（见图 10.3），因为这类混频器出现在那些专门来设计乘法器的混频器之前。

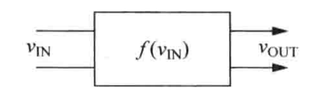


图 10.3　通用的二端口网络的非线性表示

如果非线性的表现是"理想"的（就数学意义上讲），那么我们可以用以下的级数展开式来描述输入-输出间的关系：

$$v_{OUT} = \sum_{n=0}^{N} c_n (v_{IN})^n \tag{3}$$

采用 N 阶非线性度作为混频器，信号 v_{IN} 是 RF 输入信号与本地振荡器信号之和。一般来说，输出将包括三类乘积：DC 项、输入的谐波以及这些谐波的交调乘积项[14]。并不是所有这些频谱分量都是所希望的，所以在混频器设计挑战的一部分是设计电路拓扑使其从根

---

[12] 要记住，信号的频谱近似于以 $1/n$ 速率滚落，其中 n 是从信号时域表征中产生冲击的求导次数。实际中人们感兴趣的大部分信号具有足够快的滚落特性，因而一般情况下高于 5 次或 7 次的谐波都已经消失殆尽了，但具体情况可能会有所不同。

[13] 我们很快会看到三端口混频器的优势。

[14] 要记住基本的波是那些谐波。

本上几乎不产生所不希望的项。

公式(3)中的偶次非线性系数产生了 DC 项，这些可以通过 AC 耦合隔直电容很容易地过滤掉。$m\omega_{LO}$和 $m\omega_{RF}$谐波项从基波($m=1$)一直向上延展到 N 次谐波。与 DC 项一样，它们常常也比较容易被滤掉，因为它们的频率通常距离希望得到的 IF 很远。

交调(IM)乘积项是那些各种和频与差频项。它们的频率可以表示成 $p\omega_{LO} \pm q\omega_{RF}$，这里，整数 p 和 q 大于零并且它们之和最大为 N，只有二阶交调($p=q=1$)是通常所希望得到的[15]。遗憾的是，其他 IM 积的频率也可能会很接近希望得到的 IF，因而很难将它们消除，正如我们将要看到的那样。由于一般来说高次非线性(也就是在幂级数展开式中 n 大的值)往往会产生较多的这些不希望得到的项[16]，所以，如果混频器只有一个输入端口(如图 10.3 所示)，那么它们就应当近似平方律的特性。我们现在考虑一个平方律混频器的具体性质，以说明它与高阶非线性混频器相比有什么优点。

二端口混频器例子：平方律混频器

为了了解平方律混频器中所期望的乘法来自何处，我们注意到在级数展开式中唯一的非零系数是 c_1 和 c_2 项[17]。然后，如果假设输入信号 v_{IN}是两个正弦之和：

$$v_{IN} = v_{RF}\cos(\omega_{RF}t) + v_{LO}\cos(\omega_{LO}t) \tag{4}$$

那么这个混频器的输出可以表示成三个不同分量之和：

$$v_{OUT} = v_{fund} + v_{square} + v_{cross} \tag{5}$$

式中，

$$v_{fund} = c_1[v_{RF}\cos(\omega_{RF}t) + v_{LO}\cos(\omega_{LO}t)] \tag{6}$$

$$v_{square} = c_2\{[v_{RF}\cos(\omega_{RF}t)]^2 + [v_{LO}\cos(\omega_{LO}t)]^2\} \tag{7}$$

$$v_{cross} = 2c_2 v_{RF} v_{LO}[\cos(\omega_{RF}t)][\cos(\omega_{LO}t)] \tag{8}$$

基波项就是与原始输入成比例的部分，因此代表着没有任何有用的混频器输出。它们必须通过滤波去除。同样，v_{square} 分量也代表着没有任何有用的混频器输出，因为这可以从公式(1)的以下特殊情形中明显地看出：

$$[\cos(\omega t)]^2 = \frac{1}{2}[1 + \cos 2\omega t] \tag{9}$$

因此我们看到 v_{square}分量产生了 DC 偏置以及输入信号的二次谐波。这些一般也必须通过滤波去除。

有用的输出来自 v_{cross}分量，这是因为在公式(8)中明显地有乘法。利用公式(1)可以把 v_{cross}重写成能更清楚表达混频作用的形式：

$$v_{cross} = c_2 v_{RF} v_{LO}[\cos(\omega_{RF} - \omega_{LO})t + \cos(\omega_{RF} + \omega_{LO})t] \tag{10}$$

对于固定的 LO 幅值，IF 输出幅值线性地正比于 RF 的输入幅值。也就是说，这一非线性实现了一个线性混频，因为输出正比于输入。

这一非线性的变换增益很容易从公式(10)中求出：

---

⑮ 一个给定的交调项的阶数是 p 和 q 之和，所以二阶交调乘积项来自级数展开中的二次项。

⑯ 如同大多数将特例普遍化一样，这里也有例外。例如，在设计频率乘法器时，高阶非线性是非常有用的。然而，在设计混频器时，通常是不希望出现高阶非线性的。

⑰ 这里也可能会有非零 DC 项(即 c_0 可能不为零)，但这个分量很容易通过滤波除掉，所以我们在一开始就忽略它，以便简化方程的拥挤程度。

$$G_c = \frac{c_2 v_{RF} v_{LO}}{v_{RF}} = c_2 v_{LO} \tag{11}$$

与任何其他增益参数一样，转换增益可以是一个无量纲的量(或一个跨导，跨阻等)。在分立元件设计中习惯上把变换增益表示成功率比(或它的等效分贝值)，但在典型的 IC(集成电路)混频器中，由于输入与输出阻抗大小的不同，采用电压或电流转换增益也是合适的。当然，为了避免混淆，有必要说清楚是何种类型的增益[18]。

正如前面提及的那样，平方律混频器的优点是：不希望有的频谱分量通常与中频频率相距较远，因而很容易滤掉。鉴于此，二端口混频器常设计成在最大实用范围内与平方律的特性一致。

平方律混频器几乎可以用任何常见的非线性来实现，因为二次项通常是主要的。在图 10.4 中，偏置、RF 和 LO 项作为驱动串联在三极管的基极。偏置电压 V_{BIAS} 可以根据需要选择，以使有源器件置于要求的工作区。比如，JFET 或 MESFET 需要负偏置，通常的 MOSFET 和三极管需要一个正的偏置。当然为了提供一个稳定的偏置，还需要某些改进。

输出电路中的集总并联谐振回路可以用 $\lambda/4$ 的传输线实现，这是指那些工作频率，其对应的尺寸可以被容纳。实际电路中 RF 和 LO 的相加可以用电阻和电抗的组合来实现。因为 RF 和 LO 信号是串联的，它们之间的隔离很差。另一种(但在功能上是等效的)能减少在 RF 端口上相对大的 LO 信号影响的电路如图 10.5 所示。

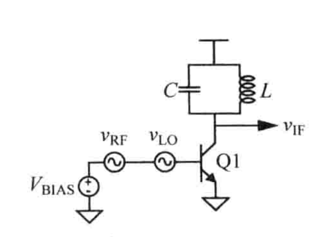


图 10.4　平方律关系混频器(简化原理图)

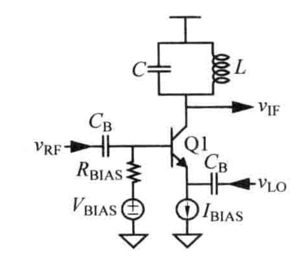


图 10.5　平方律关系混频器(另一种结构)

RF 信号直接驱动基极(通过一个隔直电容)，而 LO 则驱动发射极。采用这种方式时，基极-发射极的电压就是以地为参考的 LO 和 RF 信号之和。偏置电流直接由电流源确定，而 DC 基极电压由 V_{BIAS} 的值决定。电阻 R_{BIAS} 的值要足够大，以避免过多的负载效应，同时也使它产生的噪声影响最小。

并不一定要完全符合平方律的特性才能实现混频功能。图中的三极管电路能工作是由于当输入幅值在某一定范围内，i_C-v_{BE} 关系的级数表达式中的二次项占据支配地位。准确地说，在某个适当限制的范围内，许多非线性特性都可以用一个平方律形状来很好地近似，因此，一旦求出这个二次系数(c_2)的值，就可以估算用作混频器的其他非线性器件的转换增益。为了说明这一点，我们来估算一个双极型晶体管的转换增益。为了简化计算，继续忽略

[18] 很多时候，IC 实现公布的功率增益实际上是电压增益的测量结果，因此当输入输出阻抗严重不匹配时，电压增益会错误地反映功率增益，而大多数情况下，输入输出阻抗大不相同。所以有必要强调功率与电压的区别。

动态效应，于是可以使用指数的 v_{BE} 定律：

$$i_C \approx I_S e^{\frac{v_{BE}}{V_T}} \tag{12}$$

这里 V_T 是热电压 kT/q，而不是阈值电压。

把这个熟知的关系展开至二次项得到[19]：

$$i_C \approx I_C\left[1+\frac{v_{IN}}{V_T}+\frac{1}{2}\left(\frac{v_{IN}}{V_T}\right)^2\right] \tag{13}$$

通过观察（是的，几乎是可以这样得到），有

$$c_2 = \frac{g_m}{2V_T} \tag{14}$$

所以，转换增益可估算得到：

$$G_c = c_2 v_{LO} = g_m \cdot \frac{v_{LO}}{2V_T} \tag{15}$$

这里的变换增益就是跨导，它既正比于标准的增量跨导，也正比于本地振荡器驱动幅值与热电压之比。因此双极型晶体管的转换增益依赖于偏置电流、LO 幅值以及温度。

我们现在考虑一个理想的平方律长沟道场效应晶体管（FET），对该晶体管有

$$i_D = \frac{\mu C_{ox} W}{2L}(V_{GS}-V_T)^2 \tag{16}$$

（译者注，这里 V_T 是阈值电压，不是热电压）由于具有速率饱和效应，短沟道（高电场）器件线性度较好，因此用作混频器时一般不如长沟道器件来得好[20]。

如果栅-源电压 V_{GS} 是 RF、LO 及偏置这三项的和，那么可以写成：

$$i_D = \frac{\mu C_{ox} W}{2L}\{[V_{BIAS}+v_{RF}\cos(\omega_{RF}t)+v_{LO}\cos(\omega_{LO}t)]-V_T\}^2 \tag{17}$$

由上式可以很容易地求出转换增益（这里是跨导量）：

$$G_C = \frac{\mu C_{ox} W}{2L} v_{LO} = \frac{I_D}{V_{OD}^2} v_{LO} = g_m \frac{2v_{LO}}{V_{OD}} \tag{18}$$

其中，V_{OD} 是栅极的过驱动电压（$V_{GS}-V_T$）。注意，这个理想的平方率场效应晶体管的转换跨导增益和双极型晶体管有同样的形式，所不同的只是过驱动电压代替了热电压的作用。同时要注意，这个理想的器件的转换跨导是和偏置无关的[21]，并且也与温度（通过迁移率变化）和 LO 驱动幅度有关。

就像上面对 MOSFET 的推导，前面对双极型晶体管的计算忽略了基区和发射区的串联电阻。这些电阻可以使晶体管特性线性化，从而减弱了混频效果。因此，选择合适的器件类型对于性能最优化是很有必要的。

另一种改进 RF-LO 隔离的方法是使用双栅 FET 混频器。正如它的名字暗示的，这个混频器使用了一种拥有两个栅的 FET。对于实际的应用，这种结构可以等效成构成

---

⑲ 我们隐含假设加在基极-发射极间的驱动信号包含了 DC 成分以及 RF 和 LO 成分，所以 I_C 不为零。

⑳ 再次提醒读者，“短沟道”意味着“强电场”。因此，在适当小的栅-源电压下，即便是“短”器件也会有平方率的特性。

㉑ 这个对偏置的无关性只适用于在平方率关系成立的工作区域，因此必须提供足够的偏压来保证这一关系成立，所以 V_{BIAS} 不能为零。由于 MOSFET 在弱反型区具有指数特性，通常将 V_{BIAS} 选得足够大，以使得晶体管的栅-源电压总是高于其阈值电压。

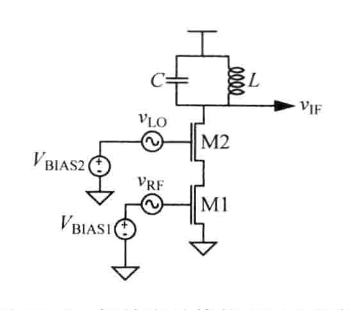


图 10.6　场效应晶体管(FET)双栅混频器(简化原理图)

cascode 电路拓扑的两个 FET，见图 10.6。

在这个电路中，M1 的作用是一个故意偏置成具有较差输出电阻的跨导器，它通常工作在三极管区。偏置这样选择了之后，加在 M2 上的 LO 驱动引起 M1 的漏极电流和电阻跟着 LO 的频率变化。这样 M1(因此也就是 M2)的漏极电流就像其他混频器一样包含和频或差频成分。

正如其他电路一样，在实际可行的频率下，LC 谐振回路可以为等价的传输线所替代。

10.4　基于乘法器的混频器

我们已经看到产生混频作用的非线性是通过它们偶然提供的乘法而实现的。确切地说，由于乘法只是间接地产生混频，所以这些非线性通常会产生许多不希望的频谱分量。此外，由于二端口混频器只有一个输入端口，所以 RF 和 LO 信号之间无法进行很好的隔离，而缺乏隔离可能引起上面提到的问题，如 IF 放大器的过载，以及 LO 信号(或它的谐波)通过天线辐射回去。

直接基于乘法器的混频器一般都具有很好的性能，因为在理想情况下它们只产生所希望的交调乘积项。而且由于乘法器的各个输入分别进入各自的端口，在这三个信号(RF、LO 和 IF)之间可以实现很好的隔离。最后，FET 工艺提供了性能很好的开关，因而可以利用开关来实现性能优异的乘法器。

10.4.1　单平衡乘法器

一类极为普遍的乘法器是先把输入的 RF 电压变换成电流，然后在电流域内执行乘法。这一类最简单的乘法单元如图 10.7 所示[22]。MOSFET 的符号仍代表任何有源器件。这种结构，加上恰当的满足不同需要的偏置电路，可以使用 MESFET，MOSFET，JFET 和双极型晶体管。

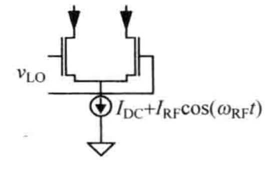


图 10.7　单平衡混频器

在这一混频器中，v_{LO} 要足够大，以使两个晶体管在 LO 频率下交替导通(交互)，把尾电流轮流从一边切换到另一边，所以尾电流实际上是与频率为 LO 频率的方波相乘[23]：

$$i_{out}(t) \approx (\mathrm{sgn}[\cos\omega_{LO}t])[I_{BIAS} + I_{RF}\cos\omega_{RF}t] \tag{19}$$

由于方波由基波的奇次谐波组成，所以尾电流乘以方波就得到了如图 10.8 所示的输出频谱。

---

[22] 通常将这一总类的混频器统称为 Gilbert 混频器，但只有部分是真正意义上的 Gilbert 混频器。真正的 Gilbert 乘法器假设所有的输入变量都已能用相应的电流形式来表示，进而完全对电流进行操作，而不涉及 *V-I* 的转换问题。见 Barrie Gilbert 的里程碑式的论文，"A Precise Four-Quadrant Multiplier with Subnanosecond Response"(一个精确的具有亚纳秒响应速度的四象限乘法器)，*IEEE J. Solid-State Circuits*，December 1968，pp. 365-373。

[23] 也可以将本振信号与射频信号的输入端口互换，但这样得到的混频器具有较低的转换增益和差的噪声性能，还有其他缺陷。

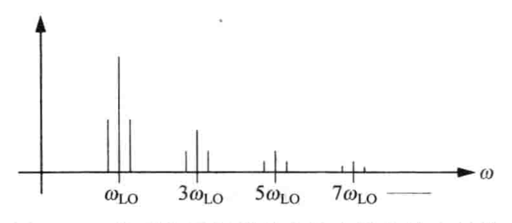


图 10.8　单平衡混频器具有代表性的输出频谱

因此输出中含有和频分量及差频分量，每个 LO 的奇次谐波都与 RF 信号混频。此外，作为 DC 偏置电流与 LO 信号相乘的结果，LO 的奇次谐波直接出现在输出中。由于在输出频谱中存在 LO，这类混频器被称为单平衡混频器。我们紧接着会研究双平衡混频器，它利用对称性抵消了不希望的 LO 输出分量。

尽管图 10.7 中的电流源包含一个严格正比于 RF 输入信号的分量，但所有实际混频器的 V-I 变换器却是非理想的。因此，一个重要的设计挑战就是最大程度地提高 RF 跨导器的线性度。无论是在共栅还是共源（或共基、共发）的跨导器电路中，常常通过某种源极负反馈来提高线性度，参见图 10.9。共栅电路采用源极电阻 R_S 使传输特性线性化。如果从晶体管源端看进去的导纳比 R_S 的电导大得多，那么这一线性化最为有效。在这一情形中，这一级电路的跨导接近 l/R_S。

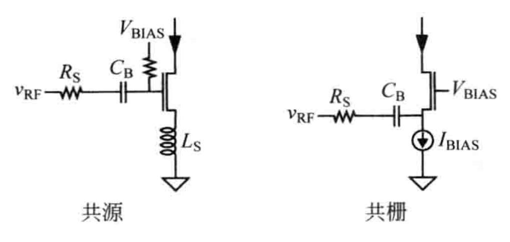


图 10.9　用于混频器的 RF 跨导器

电感简并通常比电阻简并更受欢迎，主要有以下原因[24]：电感既没有使噪声系数变差的热噪声，又没有 DC 电压降使电源有效净电压的范围减小，而这特别关系到低电压-低功耗的应用。最后，工作频率的升高使电感的电抗增加，这有助于衰减高频谐波和交调分量。

图 10.10 所示为具有线性化跨导器的比较完整的单平衡混频器。V_{BIAS}的值确定了这一单元的偏置电流，而 R_B 要足够大，从而使它不会因为负载效应影响栅电路的工作（并且也为了减小它的噪声影响）。RF 信号通过隔直电容 C_B 加到栅极上。在实际中，会采用一个滤波器从输出中去除 LO 以及其不希望的频谱分量。

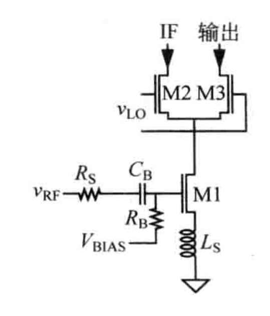


图 10.10　带有线性化跨导的单平衡混频器

通过假设由 LO 驱动的晶体管作为理想开关工作，可以估算这一混频器的转换跨导。差分输出电流可以看成是 M1 的漏电流乘以一个单位幅值方波的结果。由于一个方波基波分

[24] 有时也建议采用电容简并，但其效果显然要比电感简并差，因为电容增加了噪声和高频时的失真。由于电容不能提供直流偏置，它同时需要引入额外的偏置电路。电感元件可以解决所有这些问题。

量的幅值为 $4/\pi$ 乘以这个方波的幅值，所以我们可以写出：

$$G_C = \frac{2}{\pi} g_m \tag{20}$$

其中，g_m 是 V-I 变换器的跨导，而 G_C 本身也是一个跨导量。系数为 $2/\pi$（-3.92dB）而不是 $4/\pi$，这是因为 IF 信号被平均分成了和频分量与差频分量。

10.4.2　有源双平衡混频器

为了在一开始就防止 LO 乘积项进入到输出，可以将两个单平衡电路合在一起构成一个双平衡混频器，见图 10.11。我们再次假设 LO 驱动足够大，从而使差分对的作用像一对电流换向开关。注意，两个单平衡混频器就 LO 而言连接成反并联的，而对于 RF 信号则是并联的。因此 LO 项在输出端的和为零，而变换后的 RF 信号在输出端加倍。这样，该混频器可提供良好的 LO-IF 隔离，因而在输出端减轻了对滤波的要求。如果仔细设计，那么这一电路的 IC 实现一般都能提供 40dB 的 LO-IF 隔离，并且还有可能超过 60dB。

动态范围与在单平衡有源混频器中一样，部分受限于混频器 RF 端口的 V-I 变换器的线性度，所以大多数的设计都致力于找到更好的方式提供这一 V-I 变换。在单平衡混频器中采用的基本的线性化技术也可以应用到双平衡的情形中，如图 10.12 所示。

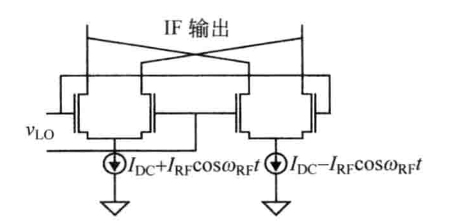


图 10.11　有源双平衡混频器

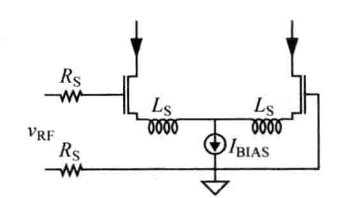


图 10.12　双平衡混频器中线性化了的差分 RF 跨导器

在低电压应用中，可以用一个并联 LC 振荡回路来代替 DC 电流源以构成一个"零净空"(zero-headroom)的 AC（交流）电流源。应当将这一振荡回路的谐振频率选择成能排除任何最不希望有的共模分量。如果有几个这样的分量存在，我们可以采用并联 LC 回路的串联组合。采用这种方式实现的一个完整的双平衡混频器如图 10.13 所示，它的变换跨导的表达式与单平衡情形时一样。

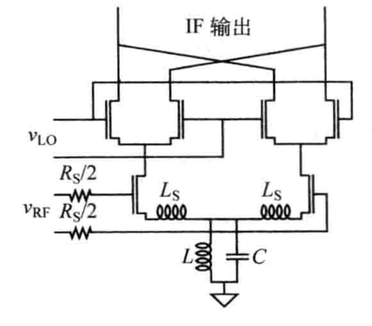


图 10.13　具有最小电源电压净空的双平衡混频器

通过在接收 RF 输入的电感性简并对上加上合适的栅极电感，可以很容易地将这些混频器改造成低噪声混频器。通过和独立的 LNA 一样的方法，可以构造出低净空、低噪声的可在一些应用中代替独立 LNA 的混频器。调节输入回路可以在转换增益、噪声系数和失真之间有可变的折中。

Gilbert 型混频器的噪声系数

计算混频器的噪声系数是困难的，因为噪声源具有周期性稳态的特性(cyclostationary)。一种基于仿真的技术包括时变脉冲响应的表征，认为混频器即使不是时不变的也至少是线

性的[25]。虽然这一方法是精确的，且十分适合于进行分析，但由于它的过度复杂性，不能提供对设计的内在理解。然而，我们依然能确定几种重要的噪声源，并且提供关于如何使噪声系数最小化的通用建议。

其中一个噪声源无疑就是跨导器本身，所以它的噪声系数就决定了混频器噪声系数的下限。在计算 LNA 噪声系数时采用的方法同样也可用于计算跨导器的噪声系数。

差分对也在好几个方面使噪声性能降低。一个对噪声系数的贡献来源于非理想的开关特性。它会引起信号电流的衰减。因此，在这样的混频器设计的一个挑战就是如何设计这些开关(以及相关的 LO 驱动)使衰减尽可能地小。

开关晶体管对噪声系数的另一个影响表现在当两个晶体管都导通电流时，此时会产生噪声。此外，在 LO 中的任何噪声也在这个有源增益区间被放大。使这个同时导通时间尽可能短可以减少这种性能的下降，所以必须要提供足够的 LO 驱动，以使这一差分对在最大的实用范围内接近理想的无穷快的开关。最后，忽略和频信号或差频信号本身就造成了 3dB 衰减，自然也会使噪声系数恶化(3dB)，因为噪声不可能如此容易地去除。结果是，实际电流模式的混频器一般至少显示出 10dB 的 SSB 噪声系数，并且这些值常在 15dB 左右。

Gilbert 型混频器的线性度

这类混频器的 IP3 受限于跨导器的 IP3，所以用来估算一般放大器 IP3 的三点法也可以在这里用来估计跨导器的 lP3。如果由 LO 驱动的晶体管能发挥性能良好的开关作用，那么整个混频器的 1P3 一般不会与跨导器的 IP3 有什么差别。为了获得好的开关特性，非常重要的是要注意(尽管必须要有足够的驱动)避免过大的 LO 驱动。为了理解过大的 LO 驱动是不利的而不是有益的，让我们来考虑总是存在的寄生电容对差分对共源连接的负载效应。由于每一个栅的驱动都远远超出了达到理想开关所必需的驱动，所以共源点电压也同样被过驱动，于是就出现了电流尖峰。在极端情形下，这一尖峰电流会使晶体管离开饱和区。即使这一点不发生，输出频谱也将由来自尖峰电流的分量而不是由降频变换后的 RF 信号占支配地位。因此，我们应当只采用足够的 LO 驱动来保证可靠的开关作用，而不能过大。

关于采用时域模拟器模拟混频器 IP3 的简短说明

如同我们在放大器交调失真的仿真中提到的，一般电路仿真器(如 SPICE)如果真想要实现精确的混频器仿真也只是很勉强能做到的。这一问题有两个基本的原因：由于在混频器中信号的动态范围很宽，不得不采用远比“正常”电路仿真要严格的数值容差，并且重要频谱分量所跨越的大频率范围也使仿真的时间很长。因此，要从瞬态仿真中得到精确的 IP3 值通常是非常有挑战性的。此外，我们提醒读者对混频器的仿真结果不可盲目相信。

因为在某些仿真工具中甚至“精确”选项与要求的 IP3 仿真相比也差了好几个数量级，我们用来缓解这些问题的一个具体方法就是使容差逐渐变小，直到仿真结果不再显著变化。特别是在 IP3 仿真中的 IM3 分量的特性是表明容差是否足够小的一个极为灵敏的标志。如果 IM3 项并不表现为+3 的斜率(在 dB 坐标上)，那么容差很可能是太大了。我们也必须保证两个输入正弦信号的幅值要足够小(也就是远小于 1dB 压缩点或者交调点对应的输入幅值)，以保证混频器的准线性工作，否则非线性的高次项将明显影响输出，从而搞乱了结

---

[25] C. D. Hull and R. G. Meyer, “A Systematic Approach to the Analysis of Noise in Mixers” (混频器噪声分析的一个系统性方法), *IEEE Trans. Circuits and Systems I*, v. 40, no. 12, December 1993, pp. 909-919.

果。在最初的设计阶段,可以对跨导电路采用三点法来估算它的IP3,而不必为进行瞬态仿真而苦恼。

另一个微妙之处是保证在瞬态仿真中有相等的时间步长,因为FFT算法一般都假定均匀采样。由于某些模拟器采用自适应的时间步长来加速收敛,所以计算FFT时就会出现假象的频谱成分问题。我们可以把时间步长设置成感兴趣的变化最快的时间间隔的非常小的一部分以保证收敛,而不采用自适应的时间步长。例如,我们也许必须采用比RF信号周期小三个数量级的时间步长。因此对于1GHz的RF输入,也许需要采用1ps的时间步长。正是由于迭代、很小的时间步长以及数值容差等问题,都使得IP3模拟进行得非常慢。幸运的是,与放大器的情形中一样,时域仿真器的替代方法也已经开发出来以解决这些问题[26]。

其他线性化技术

由于这些电流模式混频器的线性度主要由跨导器的性能控制,所以值得考虑采用其他方法来改进线性度。原理上,可以采用四种方法;预失真(predistortion)、反馈、前馈及分段逼近。这些技术可以单独使用,也可以联合起来使用。下面是这些方法中一些典型的(但不可能是全部的)例子。

预失真是把两个相互相反的非线性串联在一起,并且它与前馈的方法一样需要仔细地匹配。预失真实际上几乎被普遍采用,因为它是电流镜工作的基础原理。在电流镜中,输入电流通过某些非线性关系转变成栅-源电压,然后它又反过来产生与输入完全成正比的输出电流。预失真对于真正的Gilbert混频器的工作来说也是很基本的,一对PN结计算出输入差分电流的反双曲正切,随后一个差分对又将这一非线性电压反演至电流。

负反馈估计出来误差,然后把信号反相后加回到输入上,由此帮助抵消失真所代表的误差。只要回路传输值很大,那么失真可大幅降低。由于负反馈系统是在事后计算误差的,所以整个闭环带宽必须保持为构成该系统的元件本身的带宽的一个很小部分,否则事后的误差估算在最好情况下不会产生任何影响,而在最坏情况下可能引起不稳定。本章的串联反馈例子则是线性化高频跨导器的一种很普遍的方法。

前馈则是另一种线性化技术,它在电路系统处理信号的同时计算误差的估计值,由此规避了负反馈的带宽和稳定问题。然而此时误差计算和误差消除取决于匹配,所以,实际实现的最大失真减少程度往往明显不及用负反馈能达到的失真减少程度。前馈在高频时最令人感兴趣,而负反馈在高频时由于回路传输值不足而变得不太有效。

对跨导器电路进行前馈修正的例子是Pat Quinn的"cascomp"(带有补偿的共发共基)电路,最早用双极型晶体管实现[27]。正如图10.14所示,这一跨导电路由cascode(共发共基)差分对构成,并且在它上面又连上了一个附加的差分对。部分线性化由源极负反馈电阻R提供,但在差分对

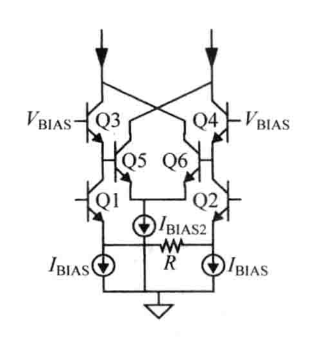


图10.14 经典的双极型晶体管cascomp(带有补偿的共发共基)电路

[26] 在实验室测量这些量时一样要小心。就如进行仿真时,两个输入正弦信号的幅值必须足够小以避免引起产生更高次谐波的非线性(这会使斜率不再是3),同时又要比噪声底限大很多。

[27] "Feedforward Amplifier"(前馈放大器),美国专利号#4146844,1979年3月27日公布,1984年重新公布。

Q1-Q2 的跨导中仍然有显著的非线性。为了看明白这一点，考虑到在电阻两端的电压是输入电压减去 Q1 和 Q2 基极至发射极电压的差：

$$V_R = v_{IN} - (v_{BE1} - v_{BE2}) = v_{IN} - \Delta v_{BE1} \tag{21}$$

由于目标是得到与 v_{IN} 严格成正比的差分输出电流，所以任何非零 Δv_{BE} 都代表误差。共发共基(cascode)连接的一对管子具有与输入对相同的 Δv_{BE}，它由里面的一对差分管度量。与这一误差成比例的电流从主电流中减去以使跨导线性化。“cascomp”的名字来自于 cascode 和“误差补偿”(error compensation)的组合。虽然为了简化起见里面的一对晶体管表示成通常的差分对，但使它线性化以降低对修正的要求常常是有利的，这样可以增加跨导保持为常数的范围。

另一种非反馈的方法是分段逼近，它是基于这样的考虑：即实际上任何一个系统在某一足够小的范围内都是线性的。这一方法把线性度的实现分配在几个系统上，每一个系统只在一个足够小的范围上工作，所以，合在一起的系统就在更宽的范围内表现为线性。

Gilbert 的双极型“多个双曲正切”电路是分段逼近的一个例子。这一电路的 MOS 形式(读者可能会犯迷糊了)如图 10.15 所示[28]。这三个差分对中的每一个分别在以 V_B，0 和 $-V_B$ 为中心的输入电压范围上具有合理的线性跨导。对于输入电压接近零的情况，跨导是由中间的差分对提供，其对于足够小的 v_{IN} 大致为常数。当输入电压明显偏离零时，尾电流最终几乎完全转向中间差分对的某一边，然而当选择合适的偏置电压 V_B 时，外部差分对中的一对将取而代之，并且继续使输出电流增加，见图 10.16。

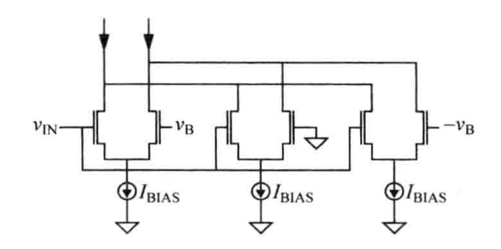


图 10.15　具有高线性度的跨导(g_m)单元

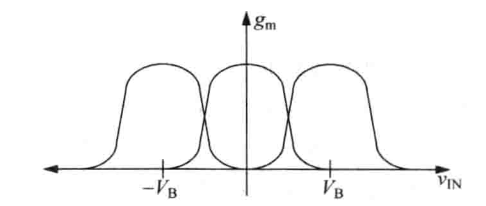


图 10.16　采用分段逼近方法的线性化方法图示

总跨导是各个分支跨导之和，通过采用足够多数量的附加差分对，并使每对恰当地排布开，就可以使这个总跨导在一个几乎任意大的范围上大致为常数。代价则是功耗和输入电容的增加。

10.4.3　无源双平衡混频器

至今我们只是研究了有源混频器，它们具有提供转换增益的特性。然而，在非常高的频率下，可能没有能提供足够增益的有源器件，所以我们也需要考虑无源混频器。除了它们有可以工作在很高频率的潜力外，这种混频器也有低功耗和工作在非常低的电源电压的潜力。

在至今所考虑的有源混频器中，RF 信号表示成电流形式而不是 RF 电压本身去有效地乘以从本地振荡器导出的方波。避免 V-I 变换问题的一种方法是直接在电压域内切换 RF 信号。这一选择用 FET 显然比用双极型晶体管更容易实现，这也就是为什么双极型混频器

[28] 这一名字源于以下事实：一个双极型差分对的传输特性是双曲正切。

几乎毫无例外地都是有源的电流模式类型。

最简单的无源换向 FET 混频器包含有四个开关连接成桥式结构(见图 10.17)。和这一章的大多数其他电路不同,双极型器件并不能很好地取代上面的场效应管器件。双极型器件是很好的电流模式的开关,但是它们不是很好的电压模式开关。

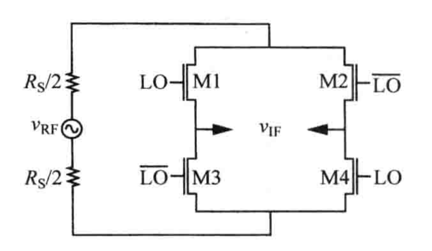


图 10.17　简单的双平衡无源 MOS 混频器

这些开关由本地振荡器信号进行反相位的驱动,所以在任何时候只有一条对角线上的一对晶体管是导通的。当 M1 和 M4 导通时,v_{IF} 等于 v_{RF},而当 M2 和 M3 导通时,v_{IF} 等于 $-v_{RF}$。与此完全等效的描述是:这一混频器把输入的 RF 信号乘以频率为本地振荡器频率的单位幅值方波。因此输出中含有许多混频的积,它们都来自方波的奇次谐波傅里叶分量[29]。幸运的是,正如前面已讨论过的那样,通常很容易将它们滤波掉。

这一基本单元的电压转换增益很容易根据前面的描述计算出来。假设乘以单位幅值的方波,可以立即写出:

$$G_c = \frac{2}{\pi} \tag{22}$$

这里 $2/\pi$ 的因子同样来自于把 IF 能量均分在和频分量与差频分量上[30]。

实际上,确切的电压转换增益可以有些不同于 $2/\pi$,因为现实的晶体管不会在零时间内切换。因此一般来说输入的 RF 信号不是乘以一个纯方波信号。然而与直观也许恰好相反,这一与理想假设的偏离通常会使电压变换增益增加到 $2/\pi$ 以上。

更为一般的电压转换增益表达式的推导有些复杂,所以我们这里只说明对问题的一些有关内在理解[31]。混频器的输出可以看成是两个时变分量与一个比例系数的乘积:

$$V_{IF}(t) = v_{RF}(t) \cdot \left[\frac{g_T(t)}{g_{Tmax}} \cdot m(t)\right] \cdot \left[\frac{g_{Tmax}}{\bar{g}_T}\right] \tag{23}$$

函数 $g_T(t)$ 是从 IF 端口看到的时变的戴维宁等效电导,而 g_{Tmax} 和 $\bar{g}_T$ 分别为 $g_T(t)$ 的最大值和平均值。混频函数 $m(t)$ 的定义为

$$m(t) = \frac{g(t) - g(t - T_{LO}/2)}{g(t) + g(t - T_{LO}/2)} \tag{24}$$

其中,$g(t)$ 是每个开关的时变电导,而 T_{LO} 是 LO 驱动的周期。混频函数没有任何 DC 分量,是周期函数,周期为 T_{LO}。如果 LO 信号是精确半波对称的,该函数只含奇次谐波分量。

公式(23)第一个方括号里项的傅里叶变换对于方波驱动(如早前所认定的)在 LO 频率分量的幅度为 $2/\pi$,而对于正弦波驱动则为 $1/2$,所以有效的混频函数对于方波驱动确实产

㉙ 然而,这种情形与电流模式的混频器相同。而且如果方波的占空比不是精确的 50%,那么 LO 项的偶次谐波也不为零。

㉚ 如果我们假设源端和负载端的负载终断情况相同,那么这一增益对应于 3.92dB 的电压和功率损失。由于许多实际的电路(如 10.4.4～6 节中讨论的分立无源混频器)还存在另外的衰减原因(如非零的开关压降、趋肤效应等),它们一般都展现出比理论极限更大一些的变换损耗。这类混频器的变换损耗通常在 5～6dB 左右。

㉛ 详细的推导过程见 A. Shahani *et al.*, "A 12 mW Wide Dynamic Range CMOS Front-End for a Portable GPS Receiver"(一个用于便携式 GPS 接收器的宽动态范围 CMOS 前端电路), *IEEE J. Solid-State Circuits*, December 1997。

生较高的变换增益。然而，第二个方括号里的项对于方波驱动（由于峰值电导与平均电导相等）为 1，而对于正弦波驱动为 $\pi/2$。因此总的转换增益在正弦波驱动时较大，这是因为第二项的补偿超过了（有效）混频函数较小的影响，然而差别并不特别大。对于正弦波驱动，变换增益为 $\pi/4$（−2.1dB），而在方波驱动时得到的增益为 $2/\pi$（−3.92dB）。

由于（有效的）混频函数频谱的原因，不希望的乘积项会出现在这类混频器的 IF 端口。因此滤波问题值得仔细考虑，特别是关系到输入和输出终断问题的时候。在分立元件设计中，信号源阻抗和负载阻抗通常是实数并且就是 50Ω。在其他情况下，比如在集成电路实现中，在 IF 端口的负载可能不是这样的终断。比如，MOS 形式，作为一个很好的近似，在 IF 端口的负载通常是容性的。这种情况下，可以利用电容负载和开关的电阻做成一个简单的低通滤波器。彻底的分析[32]显示这个滤波器的传递函数是简单的：

$$H(s) = \left[s\frac{C_L}{\bar{g}_T} + 1 \right]^{-1} \tag{25}$$

我们看到，极点频率就是平均电导（同样，是从 IF 端口通过开关向后看）与负载电容之比。这一内含的滤波作用可以被利用来提供对混频器所产生的不希望乘积项的衰减。

噪声系数和 IP3 都是 LO 驱动的强函数，因为在"导通"状态开关的电阻值必须保持在较低且恒定才能同时优化这两个参数。IP3 也与 L 匹配提供的电压提升的数量有关。可以将这一提升调小，从而用转换增益来换取 IP3 的改善，同时在某些情形中完全移去 L 匹配也许是恰当的。采用 300mV 的 LO 驱动幅值时，10dB 的典型 SSB 噪声系数和 10dBm 的输入 IP3 是很容易达到的[33]。作为一个粗略的估计，这类混频器的 DSB 噪声系数与功率转换损耗大致相同。

作为对这类混频器噪声性能的最后一个注解，我们可能会认为没有 DC 偏置电流就意味着不存在 $1/f$ 噪声。然而，因为混频器是一个周期的时变系统，因此，比如处于本地振荡器频率整数倍处的频谱分量可以向下变频而折到 DC。这样，$1/f$ 噪声仍然可以出现在混频器的输出中而不需要混频器本身有任何 DC 偏置电流。这就是说，不管怎么说 DC 电流还是出现了。当减小混频器输出的 $1/f$ 噪声是很重要时，一般采用以下几种方法会比较有效：(a)把 LO 的驱动减小到实现混频作用所需要的最小值；(b)仔细设计本地振荡器，特别是使它在有用频率附近（close-in）的相位噪声减到最小（见第 18 章）。最后，将混频器输出端接较大的电阻负载，可以使 DC 电流较小，从而减小 $1/f$ 噪声。这些考虑在设计对 $1/f$ 噪声敏感的接收机时尤为重要，其中包括直接变频（也称为零拍式或零中频）接收机以及低中频接收机。

为了减少 LO 驱动器的功耗，开关的栅电容可以与一个电感谐振（对于窄带应用来说），这可以使功率减小为原来的 $1/Q^2$。即使是在吉赫兹的频率下，把功耗降至毫瓦级或更低也是很容易的。

10.4.4 单二极管混频器

最简单也是最早的无源混频器采用单个二极管。在图 10.18 的电路中，输出 RLC 振荡

---

[32] 详细的推导过程见 A. Shahani *et al.*，"A 12 mW Wide Dynamic Range CMOS Front-End for a Portable GPS Receiver"（一个用于便携式 GPS 接收器的宽动态范围 CMOS 前端电路），*IEEE J. Solid-State Circuits*，December 1997。

[33] 这个值用于正弦波 LO 信号。

回路被调谐至所希望的 IF,而 v_{IN} 是 RF、LO 和 DC 偏置分量之和。二极管的非线性 V-I 特性使二极管电流包含许多谐波和交调频率分量,但振荡回路只选择 IF 的频率。

人们可能会认为这样一个简单的电路既不能提供任何转换增益,也不能提供任何隔离,而倾向于抛弃它。然而在极高的频率下,不可能采用任何其他的方法来实现混频器。事实上,二次世界大战采用的雷达设备中所有的探测器都是单二极管电路。没有这类混频器,许多现代毫米波频段应用就不可能发生。关于这类混频器更详细的原理图见图 10.19。

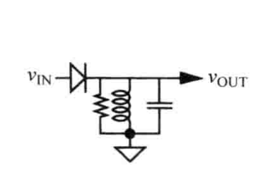


图 10.18 简单的二极管混频器

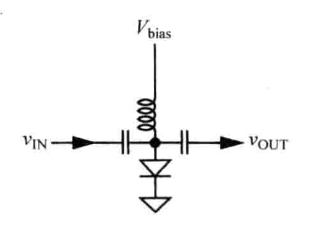


图 10.19 简单的单二极管混频器

这里,输入只是 RF 和 LO 信号之和;DC 偏置是通过一个 RF 扼流圈(或者 λ/4 的线)提供的。为了防止这一偏置影响驱动输入的电路和为输出所驱动的电路,采用了电容耦合。在输入端必须要有的求和运算可以通过各种不同的方法实现,比如电阻合成器或者各种混合型耦合器。后者由于能够提供隔离而更受欢迎。

关于这一电路的另一个注解是,如果输入信号是 AM 信号(无论是 RF 还是 IF),那么它可以用来作为 AM 信号的一个粗糙的解调器。当以这一方式使用时,可将输出电感完全除去,也不用 LO,而是用一个简单的 RC 网络来实现输出滤波,成千上万的"矿石"收音机都采用过这类检测器(这里也可称为包络检测器),甚至在今天制造的大多数 AM 超外差式收音机也采用单二极管解调器。

10.4.5 双二极管混频器

看起来好像可以有无数多种的方式来采用二极管实现混频器。正如我们将要看到的,似乎一个二极管桥式电路可以用作几乎任何一种用途,这取决于哪些端口定义为输入和输出以及二极管的朝向如何[34]。

采用两个二极管就可以构成一个单平衡混频器。在这一情形中,我们可以在 LO 和 IF 之间实现隔离,但 RF-IF 的隔离较差,如图 10.20 所示。这种基于变压器的版本在低频和中等频率工作的很好(可以说,能够到 1～2GHz)。在这个频率范围内,很多商用的变压器的实现方式是用三股线绕的圆形线圈。其中两股相互串联,用于实现次级线圈。这种实现方法保证了次级线圈良好的对称性和初级线圈与次级线圈之间良好的耦合。

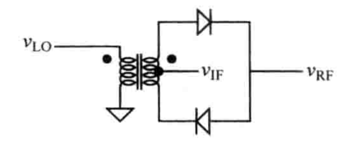


图 10.20 经典的基于变压器的单平衡二极管混频器

[34] 二极管的非线性电容可以被用来构建参量放大器。这一非线性可以用来把能量从一个本机振荡器(在参量放大器的术语中称为"泵")传送到信号中,而不是较为传统地把功率从 DC 电源传送到信号频率。参量放大器的噪声可以很低,因为只需纯电抗就可以使其工作。在微波激射器(maser)发明之前,这种放大器在微波频段的噪声是最低的,使得射电天文学用此取得了重大进步。在第 9 章中,对参量放大器有略微更详细的讨论。

假设 LO 的驱动足以使二极管作为开关工作而不管 RF 输入的数值有多大，当 v_{LO} 为正值时，两个二极管都导通(注意在变压器绕组上的同极性标记点)，于是把 v_{RF} 有效地连接到 IF 输出端。当 v_{LO} 变为负值时，二极管开路，使 v_{RF} 断开。

无论何时，只要二极管导通，就会将 RF 和 IF 端口连在一起，因此，RF 和 IF 之间隔离是差的这一点也是不言而喻的。同样，应当很清楚，对称性保证了 RF 和 LO 之间有很好的隔离。无论何时，当二极管导通时，RF 电压只能在变压器的绕组上产生共模电压，所以在 LO 端口上不会感应出任何电压(至少在理想情况下是这样的，实际中的非对称性总是会破坏这个完美的情况)。

在高频条件下，变压器几乎总是被某种类型的混合型耦合器替代。一个典型的例子使用了一个环形混合型耦合器来提供反相的二极管驱动，见图 10.21。为了使这种配置提供良好的性能，LO 和 RF 信号必须都在混合型耦合器的窄的带宽内。因此，最大允许的频率间隔是 10%。幸运的是，这个限制并不妨碍合理的频率计划。另一个要求是 IF 端口在 RF(和 LO)输入信号的频率处呈现出一个非常低阻的负载。这可以采用集总谐振电路(在低频时)或者传输线段(在较高频率时，如图 10.21 所示)来实现所需的电学行为。我们必须确保元件的寄生不会引起滤波器去衰减所希望的 IF 信号。第二个要求是 IF 端口在所需的(比如，差频)IF 频率下反射系数是小的，这对镜像(比如，和频)IF 频率也一样。常见的疏忽是没对镜像频率分量接匹配终端，这可能会导致混频器出现病态行为[35]。

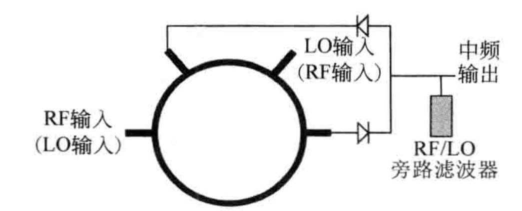


图 10.21 采用鼠奔混合型变换器的典型单平衡二极管混频器(滤波器未画出)

注意，正如图 10.21 所表明的，输入端口可以用两种方式驱动。对于这两种可能的驱动连接方式，转换损耗是相同的，隔离度也与选择哪种方式无关。但是，由这两种选择产生的杂散信号并不相同。取决于系统的细节，一种连接方式可能优于另一种。

连接混合型耦合器和二极管的线也需要长度的匹配(更准确地说，它们带来的相位差必须很小)。如果没能很好地满足这个要求，转换损耗就会增加，匹配变差，线性度也会受影响。这根线在连到二极管的整个距离上要尽可能完全是传输线。更进一步讲，应当选择寄生效应与工作频率相适应的二极管封装。由于最好的封装是没有封装，在高频下常采用硬引线作为端口的器件。

一个更为微妙的考虑是，对于 IF 信号，二极管的馈线需要加以改造以对中频频率形成交流接地的路径。在高的频率下，使用一个高特征阻抗的在 RF(LO)频率下为 $\lambda/4$ 的终端接地的传输线更为好些。在这里需要两根这样的传输线段，从而将每一个二极管到地。如果二极管需要工作在某个 DC 偏置下(偶然会有这样的情形，但不经常)，这个传输线段可能

[35] 这些病态行为并不总是不希望的，但要依赖它们以达到某种性能是十分危险的做法。

接到 DC 偏置电源上。在 RF/LO 的频率处，这个连接必须做到很好的旁路接地。

关于单平衡情形最后要说的是，应该注意在混合型耦合器内部的面积足够大以允许包含所需的电路，同时要十分小心来避免不希望的耦合。一个恰当的方法是把二极管和其他元件放置在离环至少 4～5 个电介质厚度的距离之外。通常，环本身也会变形为矩形或椭圆形，以适应电路版图的约束，如图 10.22 所示。这样做的一个不足之处是：为了验证设计功能是我们所要求的，电磁场仿真的工作量必然会增大。

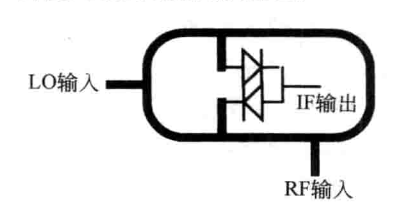


图 10.22 带有非圆环形的环形混合型变换器的混频器版图（IF 输出和 RF/LO 旁路电路未画出）

不要在家里（或任何地方）尝试这个

为了与 Horowitz 和 Hill 十分受欢迎的 *The Art of Electronics* 一书中"坏的电路主意"一节遥相呼应，我们这里提一下，单平衡混频器一个非常普遍的变种用分是分支线耦合器取代环形混合型耦合器，而后者只提供了正交的相移而不是反相，经过这样改变的结构依然能工作，已经是十分令人吃惊了。更奇怪的是这种结构还很流行。用接受严重的性能损失来换取半个波长的互连线的节省，看起来是一个不合算的折中。使用这个近似的另一个理由可能是因为这是一个宽带分支线耦合器。但是，一个性能并不好的宽带混频器如何被使用倒是个问题。

10.4.6 双平衡二极管混频器

通过再增加两个二极管和一个变压器，我们就可以构成一个双平衡混频器来提供在所有端口间的隔离（如图 10.23）。我们先给出一个采用传统变压器的实现，但是这个结构也可以被转换成准平面的形式，比如 Marchand 巴伦，就如我们稍后会看到的[36]。

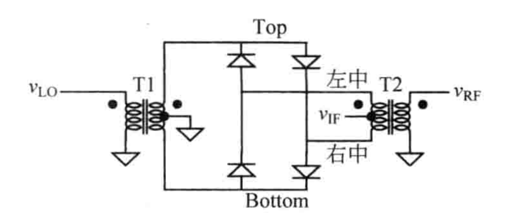


图 10.23 经典的双平衡二极管环型混频器

正如基于变压器的单平衡版本，这个混频器一直到低于吉赫兹范围内都很流行。为了理解它是如何工作的，我们再次假设 LO 的驱动大到足以使二极管作为开关工作。在电路图中，当 LO 的驱动是负的时候，左侧的一对二极管导通；当 LO 的驱动是正的时候，右侧的一对二极管导通。

---

[36] 参见 Stephen Maas，"Harmonic Balance Simulation Speeds RF Mixer Design"（采用谐波平衡法来加速射频混频器的设计）（http://www.planetanalog.com/story/OEG20020328S0107），2002 年 3 月 28 日．也可参考他的书《微波混频器》第二版，Artech House，Norwood，MA，1993 年。Maas 已经在这个研究课题上写了很多著作。对微波混频器感兴趣的读者通过阅读他的著作会受益匪浅。

当 LO 驱动为正时，根据对称性在图 10.23“右中”(Right Mid)处的电压必定为零，因为输入变压器的中心抽头是接地，于是 v_{IF} 等于 v_{RF}。(再次注意变压器的极性标记)当 LO 驱动为负时，在“左中”(Left Mid)处的电压为零，于是 v_{IF} 等于 $-v_{RF}$。因此，这一混频器有效地把 v_{RF} 乘以一个频率为 LO 的单位幅值方波。

隔离是通过电路的对称性来保证的。正如前面提到的那样，LO 驱动使输出变压器的顶端或底端的电位为零。如果 RF 输入为零，那么就没有任何 IF 输出，因此该电路提供了 LO-IF 之间的隔离。同样，我们可以通过考虑 IF 输入为零时的情形来说明 LO-RF 间的隔离。再次要说明的是，由于在输出变压器的顶端或底端处的电位为零，所以将不会有任何初级电压，因而也就不会有任何次级电压。

假设 IF 端口看到的阻抗是驱动 RF 端口的源阻抗是一个很好的近似——只要 LO 的驱动足够大，使二极管工作在完全导通的状态，并且假设 RF 端口和 T2 的任何两个相邻的抽头之间的匝数比为 1∶1。

市场上已经有了采用分立元件实现的这些无源混频器，并且工作得非常好。它的动态范围上限一般受二极管击穿电压的限制，而隔离则与所达到的匹配程度有关(包括二极管特性和变压器绕组寄生)。

假设 LO 驱动足够大，注意到 RF 信号等效地被乘以近似单位幅值的方波信号，双平衡无源混频器的转换增益可以很容易地计算出来。基波的幅度是 $4/\pi$，乘法导致了各占一半幅度的和频与差频成分。因此，理想的转换增益是：

$$20\log(2/\pi) \approx -3.92\text{dB} \tag{26}$$

实际上，采用四个二极管组合时，当 LO 驱动大于 5～7dBm 时，混频器通常能达到的转换损失在 5～6dB 附近。这个与理想损耗的差别是由于二极管和变压器的损耗，还有就是 LO 驱动没有使二极管同时导通和关断。如果 LO 的驱动功率是受限的，混频器仍然能够工作，但是性能会下降。比如，0dBm 的 LO 功率一般产生 9dB 的转换损耗。这种混频器的压缩点(折算到输入端的计量法)跟踪 LO 的功率大小，并且通常在 LO 功率的 6dB 以下。降低的 LO 功率因此会迅速地降低 SFDR(赝空动态范围)，因为噪声底限提高了(由于衰减的原因)，同时截距点趋于消失。

如果在图 10.23 中每个二极管的位置上采用几个二极管串联，那么还可以承受更高的 RF 电平，其缺点是提高了对 LO 驱动的要求，以保证二极管的开关操作。有的混频器的 1dB 压缩点高达 21dBm，但却需要 27dBm 的 LO 驱动！

对于这些高电平混频器来说，RF 和 LO 端口的隔离变得越来越重要，但是却很少超过 30dB。在细致的设计中，要考虑 LO 信号反向回馈到 RF 端口并通过天线重新发射出去的潜在可能。可以采用滤波器或单向放大器以及合理的频率方案一起来减轻再辐射的问题。

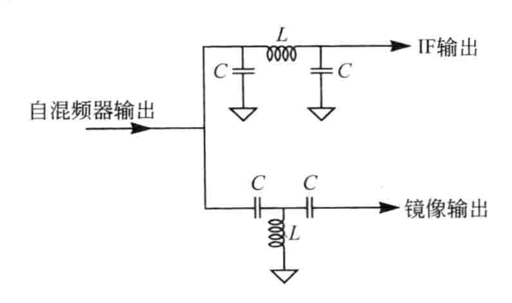
10.4.7 在镜像频率处的终断

当实际使用这些混频器时，应当注意让每一个端口接合适的特征阻抗，这是很重要的：并不只在 RF，IF 和需要的 LO 频率，也在镜像的频率。如果只使用了窄带的终端，可能反射很多交叉调制产生的分量而严重降低性能(至少会导致难以预测)。因此，只用一个标准

的 RLC 谐振回路作为带通滤波器而不用双工器[37]或是中间的缓冲级来在和频或差频产生合适的电阻负载通常是不够的。不能满足这个条件可能是很多令人费解的现象的根源。通常,不好的镜像终端将导致比预期降低的 SFDR,但是偶尔也可能改善 SFDR。为了避免不可预测的行为,关键是要注意使所有频率都接到合适的负载。

从图 10.24 的原理图中显然可见,一个简单的双工器由一个低通支路和一个高通支路组成。假设需要的 IF 是差频分量,低通支路提供了需要的输出而高通支路提供了镜像信号。使这两个终端都接合适的负载保证混频器的输出 IF 端口看到一个宽带的电阻终端。图 10.24 所示的这个独特的双工器是相对详尽的。如果对滤波的要求比较宽松,那么可以在每一个支路用一对 LC 来简单地实现。

在较高的频率,可能没有性能足够好的变压器。这种情况下,各种不同的带状线耦合器可以作为替代。特别是一个基于 Marchand 巴伦已经成功地用于从较低频率到毫米波频段的宽带混频器。通过边耦合的结构不可能实现所需的紧耦合,因此排除了微带线实现方式。而用带状线实现的宽边耦合版本要更实际。为了使图形简单,我们会用一种并不明确的反映带状线布局的方式来绘制它们,请读者在看图时注意这一点。



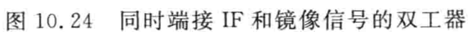


图 10.24　同时端接 IF 和镜像信号的双工器

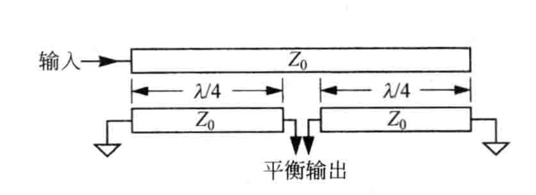


图 10.25　经典的 Marchand 巴伦

然而,经典的 Marchand(图 10.25)本身并不合适,没有足够的端口。这个小问题可以通过将上面的线切成两段而轻松的解决。这样,最左端和最右端的端口可以当做 RF(LO)和 LO(RF)输入端口,在上面的线中间新产生的端口可以用来驱动四个二极管(图 10.26)[38]。从图中我们可以看出,左边的一对耦合线给四个二极管的一对反向的端口提供了平衡的驱动(对于 RF 来说),而右侧的一对耦合线给另一对提供了一个平衡的驱动。这样,二极管看到的驱动分布和传统的基于变压器的版本完全一致。

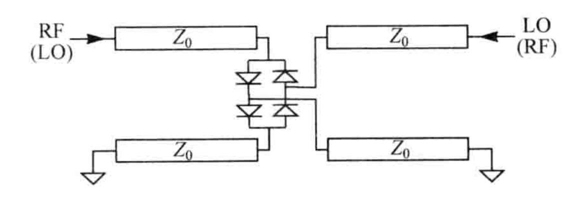


图 10.26　基于 Marchand 巴伦的双平衡混频器(只画出了核心部分,详见正文)

---

[37] 回顾一下双工器将输入信号分成高频带和低频带。双工器(diplexer)也可用作天线共用器(duplexer)的一部分(一种可实现同时收发信号的器件)。这两个名词非常相似,也经常混淆。

[38] 参见脚注 36 中引用的 Maas 的文献。

一个需要动点脑筋的部分是如何安排 IF 信号来使得它由这个结构中引出去。通过考查图 10.23 经典的结构，可以得到这样的结论：需要合成四个二极管的两个 RF 端口之间的等效中间抽头。权宜之计是在每个端口处连一根在 RF-LO 频带中心的 $\lambda/4$ 线，并把另一端接在一起。但是，IF 信号在 RF 端口也会出现。为了防止隔离度的降低，可以在 RF 驱动和四个二极管之间用电容耦合。如果电容足够小，它会衰减 IF 信号而不怎么影响 RF 信号。用一个并联谐振回路（调谐到了 IF）代替这个电容将显著改善 RF-IF 之间的隔离，但实现的难度将更大。

我们还需要为 IF 信号提供回流路径。与单平衡的情形相同，通过将四个二极管的另外两端连接一对 $\lambda/4$ 的线就能简单地解决这个问题。

还有另一种方法，既不采用隔断 IF 的电容，也不采用谐振回路。取而代之的是，RF 信号通过双工器的高通端口驱动这个结构。双工器的共模端口驱动混频器的核心，IF 信号从双工器的低通端口引出。

驱动混频器核心的巴伦以及连接到它的各种线都需要仔细的设计，以使其能够在一个很宽的频带内正常工作。耦合线的长度选取为工作频带中心处的 $\lambda/4$。这样，就可能实现在中心频率上下工作于 2～3∶1 的比例。因此，耦合器总的频率范围可以从 5∶1 到 10∶1。

但是，因为采用了 $\lambda/4$ 的传输线段，这个电路的带宽可能考虑的比较少。比如，被短路的传输线段在两倍标称频率处将呈现短路而不是开路，因此，2∶1 的频率范围成为极限。然而，如果传输线段用非常细的线（非常高的阻抗）来实现，高频损耗的增加将会阻止短路的出现，因此带宽可能要比基于无损线假设的情形要宽。

10.4.8 其他混频器结构

复数混频器

在很多调制类型中，必须要有效地跟踪镜像频率。在一些情形中，目的是为了消除不希望的镜像频率，以降低对额外的镜像抑制滤波器的要求。在其他的情形下，镜像频率只是另一个调制边带。正交（或者是复数）混频器是这种接收机的核心。

图 10.27 中，正交耦合器（通常是一个 Lange 或者分支线耦合器）将输入的 RF 信号分成同相分量和正交分量。这两个信号进入一对由同相的 LO 信号驱动的混频器。然后将输出滤波，留下的 IF 分量被送到另一个正交的耦合器，它将正交 IF 分量移相、求和以及求差。上下镜像信号（边带）可分别输出。

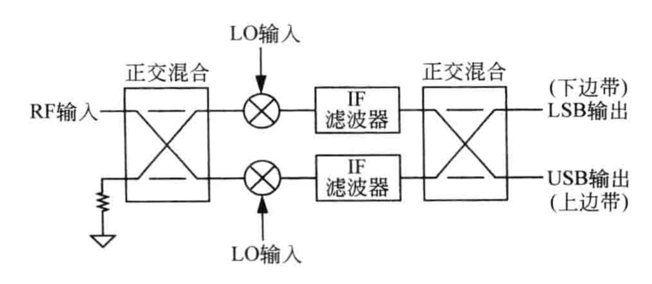


图 10.27　复数混频器示例

亚谐波混频器

在极端频率下，要产生一般超外差接收级所需的 LO 是不实际的，或者是不可能的。这种情况在射电天文学中经常出现，比如，当要探测亚毫米(或更短)波长的辐射时。我们不需要解决 RF 信号产生的问题，因为信号是由自然产生的，但是给我们留下了一个如何产生 LO 的难题。

一种解决方案是利用非线性。毕竟，非线性是很难避免的，因而可加以利用。比如，假设我们需要一个频率为 f_1 的 LO 驱动，但是并不实际或不方便得到。可以使用一个非线性元件产生那个频率的谐波，通过滤波选择需要的谐波，并且用该信号驱动混频器。

另一种选择是利用混频电路中实际产生的杂散模式。这种情况下，我们用频率 f_1/N 驱动 LO 端口，并且根据 LO 的 N 次谐波选择杂散模式。因为用来驱动 LO 端口的信号的基波是实际所需 LO 的亚谐波，这样的混频器称为亚谐波混频器。

通常混频器设计的目的是要抑制杂散模式，但是我们希望增强这个效果——至少对于一个特定的杂散模式是这样的。采用这种方式实现的一个通用的电路如图 10.28 所示。这里假定 RF 输入信号和需要的 LO 信号非常接近，以至于可以认为相等。更进一步说，假定 LO 实际提供的是它的一半的频率，所以亚谐波的阶数 $N=2$。这样，在电路上面的开路传输线对于希望得到的 RF 信号来说是开路的，但对于提供的 LO 驱动是短路的，因为在该频率处传输线的长度只有四分之一波长。同时，短路传输线对于 RF 信号是短路，而对于提供的 LO 驱动是开路。因此，二极管上的电压是 RF 信号与 LO 驱动之和。使用两个二极管反向并联保证了偶模杂散的加强。低通 IF 滤波器会尽可能去掉不想要的频率分量，但它的这种能力恰好受到该结构固有的杂散模(spur-mode)增强的限制。然而，在极高频率下，这种电路通常是唯一的选择。

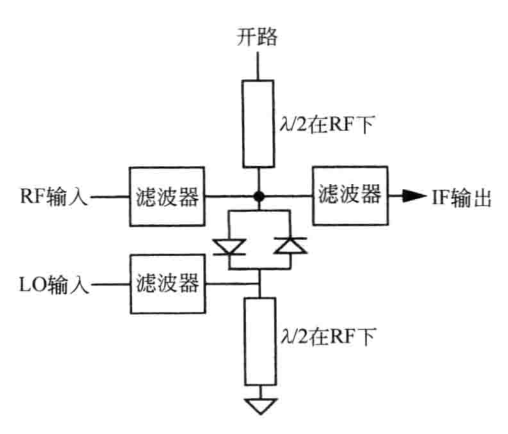


图 10.28　亚谐波混频器示例($N=2$)

第11章 晶体管

11.1 历史与回顾

随着半导体器件制造工艺日趋成熟，适合在微波频段使用的晶体管的数量和种类迅速增加。曾经一段时期，可供射频工程师选择的只有双极型晶体管(BJT)和结型场效应管(JFET)。从那以后，器件选择的种类迅速发展，包括MOSFET，VMOS，UMOS，LDMOS，MESFET，赝配和渐变HEMT(MODFET，调制掺杂场效应晶体管)，以及HBT，所有这些器件都是由不断扩充的各种材料体系来提供支持的。我们将从解释每种晶体管缩写的含义入手，尝试对这些器件进行介绍。然后将集中精力来讨论其中部分器件的建模问题。

双极型晶体管是由贝尔实验室的John Bardeen和Walter Brattain在1947年12月共同发现(是发现，不是发明)的，当时两人正按照课题负责人William Shockley[①]的要求试图构造一个MOSFET。多次实验失败后，他们开始怀疑问题出在材料的表面上，在那里衬底的严格周期性突然中断，留下没有被饱和的化学键与外来的沾污物发生结合。为了验证这一"表面态"假说，他们对半导体的表面现象进行了细致的研究。在设计来调制假设存在的表面态的一次实验中，却意外发现器件呈现出功率增益。这个器件并不是他们一直在努力构建的MOSFET，而是一个锗点接触双极型晶体管。尽管他们从未能定性地理解这个器件的行为，指望这一器件特性能重复也不现实。但这个发现足以让这个小组，包括Shockley在内，获得了1956年的诺贝尔物理学奖。虽然点接触晶体管没有取得商业上的成功，但它至少为晶体管的各个电极留下了名称。从图11.1中你可以领会基极名称的来源——它的确是点接触器件的支撑基底。

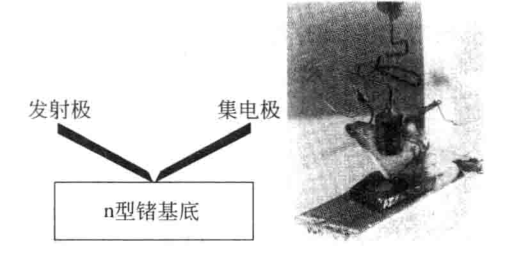


图11.1 点接触双极型晶体管(照片和版权均属朗讯贝尔实验室)

然而Shockley并不喜欢点接触型晶体管，一方面是因为这是个复杂的三维结构，另一方面是因为他自己并不是直接的发明者之一[②]。他悄悄地奋力开发一个可以被认作是自己提出的晶体管，最终他发明了今天为我们所知道的结型双极型晶体管：一个有三层极性交互变换半导体的三明治结构，而且其基本行为可通过相对简单的一维分析得以理解。与之前

---

① Michael Riordan and Lillian Hoddeson, Crystal Fire(晶体之火), Norton, New York, 1997. 对于即使只对半导体历史感兴趣的任何读者而言，这也是一本值得一读的好书。参见W. Brinkman等人写的精彩回顾文章，"The History of the Transistor and Where It Will Lead Us", IEEE J. Solid-State Circuits, v. 32, no. 12, December 1997, pp. 1858-1865.

② Shockley自己在"The Path to the Conception of the Junction Transistor"(IEEE Trans. Electron Devices, v. 23, no. 7, July 1976, pp. 597-620)中是这么说的。

发现的点接触结构不同，结型晶体管算得上是一个真正的发明，称得上是一个科学上的杰作。Shockley 于 1948 年就预言了结型晶体管的基本电学特性，这比 20 世纪 50 年代中期才实现的第一个相当粗糙的晶体管要早得多[③]。现代的双极型器件要比图 11.2 中所描绘的复杂得多。稍后当我们推导双极晶体管模型时，将会考虑相对这个图所示的更为完整的器件结构。

场效应晶体管(FET)最终于 1953 年面世，但是是以结型场效应管(JFET)的形式，而非 MOSFET 的形式出现[④]。参见图 11.3。Shockley 是在这之前一年研制出这种器件的。通过改变在反向偏置的栅结上的电压可以改变耗尽层的宽度，从而改变了一个半导体条的有效横截面，因此也就改变了流经这个条(即沟道)的电流。采用通常的制作方法，JFET 是耗尽型器件，这意味着它在常态下是导通的。来关断晶体管，必须外加偏置。由于栅电极是反向偏置二极管的一个端子，所以 JFET 呈现出很大的直流电阻，相应的其低频功率增益也就很大。

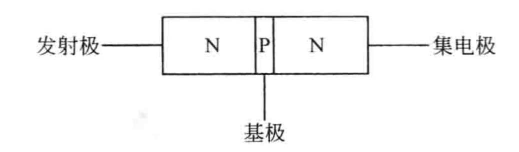


图 11.2　结型双极型晶体管(这是被大为简化后的示意图，但仍与最初的原型相似)

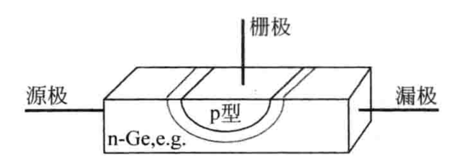


图 11.3　n 沟道结型场效应晶体管，JFET(简化了的情形，大多数实际的器件都有两个栅扩散区)

最早的晶体管是由锗制成的，但是锗的漏电流较大，这导致人们转向硅半导体。不幸的是，由于硅的熔点太高以致于生产高品质的单晶锭几乎不可能。德州仪器(Texas Instruments)的 Gordon Teal 于 1954 年初通过采用生长结的工艺成功制作出第一批硅单晶材料[⑤]。Teal 的方法使得德州仪器在硅晶体管领域开始了长达四年的垄断，直到 1958 年仙童半导体(Fairchild Semiconductor)发布了优质的硅台面器件(这种器件使用了由扩散定义的几个层)。仙童在随后的 1959 年(摩尔定律元年)又推出了平面硅器件[⑥]。

MOSFET 的研究还在继续，贝尔实验室的 Dawon Kahng 和 Martin Atalla 最终于 1960 年成功地研制出一个可以工作的硅基 MOSFET[⑦]。他们利用了这样一个惊奇的发现，即硅

---

③ William Shockley, Morgan Sparks, and Gordon K. Teal, "P-N Junction Transistors", *Phys. Rev.*, v. 83, 1 July 1951, pp. 151-162.

④ George C. Dacey and Ian M. Ross, "Unipolar Field-Effect Transistor", *Proc. IRE*, August 1953, pp. 970-979.

⑤ Teal 是半导体时代早期的无名英雄。他选择了锗作为他的博士论文研究方向。这部分是因为(如他在后来写道)"它看上去一点儿用处也没有，但正是这点吸引了我并让我觉得很有挑战"(参见 1976 年 7 月期 *IEEE Transactions on Electron Devices*)。在贝尔实验室工作时，Teal 固执地坚持单晶材料的研发，尽管这最初遭到了 Shockley 的反对，因为 Shockley 认为多晶材料永远是够用了。如果没有 Teal 研制出的具有均匀品质的高质量的单晶体，我们看到的半导体行业就不会是现在这个样子了。

由于家庭原因，Teal 离开了贝尔实验室，回到了家乡德克萨斯州。这就是他为什么会在 1954 年 5 月 10 号在美国无线电工程师学会举办的全国机载电子学会议上发表一篇论文。在他发言之前连续有好几个报告人都断言短期内不可能做成硅器件。当 Teal 从口袋里掏出几个硅基晶体管，并宣布这种管子即将由 TI 推出时，赢得全场轰动。

⑥ J. A. Hoerni, "Planar Silicon Transistors and Diodes", *IRE Electron Devices Meeting*, October 1960.

⑦ D. Kahng and M. M. Atalla, *IRE Solid-State Devices Research Conference*, Carnegie Institute of Technology, 1960. 另见 D. Kahng，美国专利号 #3,102,230，1960 年登记，1963 年授予.

本身的氧化物能极好地抑制表面态问题，而这些表面态问题在十几年前导致了 Bardeen 和 Brattain 发现了晶体管。尽管这个成功，但 MOSFET 相对较差的器件物理特性限制了它们只被用在优先考虑低成本和高集成度而非性能的场合。直到 20 世纪 80 年代后期，摩尔定律所预期的器件尺度不断缩小（因而性能不断提高，译者注）的规律才最终发力到有了微波 MOSFET 的程度。同时，由于 MOSFET 物理结构的特点，将其制作成或者是耗尽型、或者是增强型工作的器件是相对比较容易与简单的。

正如我们只简单描绘了 BJT，图 11.4 的画法也省略了 MOSFET 的大量细节。与 Kahng 和 Atalla 在 1960 年做出来的 MOSFET 相比，如今的 MOSFET 已是非常复杂的了，很难看出是由 1960 年 Kahng-Atalla 器件发展而来。这个演变的驱动力来自于对器件尺寸减小的不断追求，而同时又要求器件在面对日益增加的应力（电气的、热的与机械的）情况下仍能保证足够的可靠性。在稍后开始讨论建模这一课题时，我们将给出 MOSFET 较为更详细的结构图。

一种 MOSFET 的变体，即被称为 LDMOS(laterally diffused MOS，横向扩散 MOS)的场效应器件，被广泛应用于蜂窝通信基站的功率放大器中。LDMOS 综合恰当的增益、良好的线性度、中等大小的输出功率，以及低成本诸多优点，使得它在上面的应用中极具竞争力，参见图 11.5。就如其名字所隐喻的，LDMOS 采用了一个额外的扩散（或注入）层。增加这个 n 型层的目的本质是提供一个大的串联电阻来减小漏极附近的电场强度峰值。漏区的耗尽层延伸到此区域（而非与源区发生串通），这样就使得该器件能工作在比没有这个扩散层的器件更高的供电电压下。为此付出的代价则是导致一个较大的导通电阻。高达 70～100V 的器件击穿电压能保证其在 28～50V 的供电电压下可靠运行。与此同时，晶体管的有效栅长由 n+源扩散区与 p 阱的相对位置给出。归功于这个结构，栅长就能明显小于通常光刻机所支持的分辨率（注意在图 11.5 中，栅电极延伸到了漏区）。用在基站中的典型的 LDMOS 器件，其有效栅长在 0.25μm 数量级，这能保证器件直至 3～5 GHz 的良好性能。

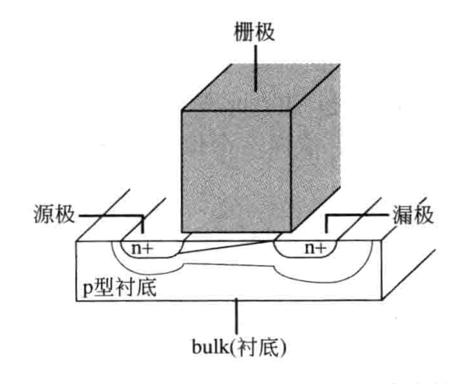


图 11.4　典型的 MOSFET（经过了大大的简化，但实际栅电极尺寸的相对大小在图中得以反映）

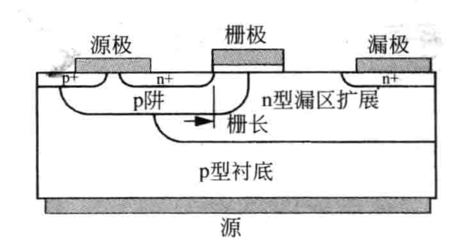


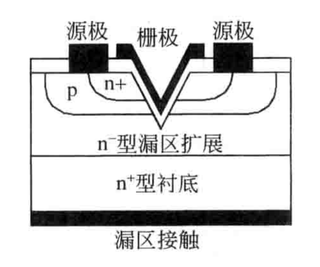
图 11.5　典型的 LDMOS 截面图（各层未按比例绘制）

为了简化封装，源极接触会做在芯片的底部。在这种情形下，通常采用一种深扩散的 p 型重掺杂区来为顶层（实际源端所在处）与芯片底部的接触之间提供低阻连接。

LDMOS 吸引人的一个原因是它能通过十分简单的设备来制造，因为它的结构只与传

统的小功率 MOSFET 稍有不同。然而，由于 LDMOS 中的电流仍沿着表面流动，它仍然是一个横向器件，相比之下，BJT 中的电流是纵向流动的。由于这样的结构单位表面积的电流密度不是很大，功率 MOS 器件会占用大量的芯片面积。

一个不大的改动可以在保证所需电流的条件下减小总的器件面积。根据晶体管的具体几何形状不同，这些晶体管被称为 VMOS 或者 UMOS 功率 FET。从图 11.6 中可以明显看出，电流主要是在垂直方向流动。事实上，对于同样数目的工艺步骤，VMOS 和 UMOS 器件能提供双倍的沟道数。此外，同 LDMOS 一样，它们的沟道长度也不由光刻精度决定。栅长是通过调节 n+源区在 p 区中的扩散深度来控制的。在光刻机的分辨率还是几微米的时代，通过这种控制可使 VMOS 获得亚微米级精度的栅长。的确，VMOS 器件是第一批商用的 MOSFET，其电特性是由载流子速度饱和作为一阶效应决定的。VMOS 和 UMOS 器件的短沟道(这导致高速工作)、电场吸纳式的漏区延伸(这可以提供高击穿电压)及垂直方向的电流流动(这使得单位芯片面积有高的电流密度)，这三者的结合使得它们在许多射频功率应用中被广泛使用。



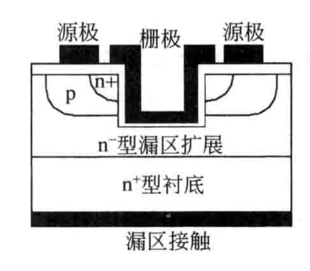


图 11.6 VMOS 和 UMOS 的截面图(各层未按比例绘制)

硅之所以在工业应用中占据主导地位是因为它易于被理解(硅是一种单元素半导体)，制造工艺简单、容易形成质量极佳的绝缘氧化物，尽管它的单一个电学参数并不突出。比如说，锗的空穴迁移率远高于硅。遗憾的是，锗的氧化物遇水就溶解了。这个问题，加之锗的高漏电流(这是源于其较低的带隙)，使得它不能成为商业上重要的半导体。

早在 20 世纪 50 年代中期，GaAs 好得多的低电场电子迁移率就备受追捧，但是掌握为生长二元半导体所需的液相外延技术的困难制约了 GaAs 原材料的得到。在 20 世纪 50 年代末期，点接触 GaAs 二极管在比 Ka 频段(Ka band)高不少的频率范围内已制备出来，但要做出晶体管则完全是另外一件事。通过掺杂来制造 BJT 已被证明比用硅更为棘手(对 GaAs 进行 p 型掺杂尤其困难)，缺少天然态的氧化物使得制作 GaAs MOSFET 相比于硅就如另一个天地那么困难。在 1965 年，加州理工学院的 Carver Mead 最终成功地制作出一个粗糙的 GaAs 晶体管，这比第一批基于 GaAs 的 LED 面世晚了 3 年[8]。通过使用肖特基(金属-半导体)二极管来形成栅极，金属-半导体 FET(MESFET)避开了 MOSFET 中的栅氧生成问题以及 JFET 和 BJT 中的通常结的形成问题。不同于现代硅工艺中的 MOSFET，MESFET 用真正的金属来构造栅极(比如覆金的 TiWN)，使得同样尺寸下 MESFET 上的串联寄生栅电阻远小于 MOSFET 的。和普通的 JFET 一样，MESFET 通常是耗尽型的(也

⑧ 参见 "Schottky Barrier Gate Field Effect Transistor", *Proc. IEEE*, v. 54, 1966, p. 307. W. W. Hooper and W. I. Lehrer，描述了一个更为实用的 MESFET，参见 *Proc. IEEE*, v. 55, 1967, p. 1237.

有增强型的，但不常见）。经过相当大的改进，GaAs MESFET 在 20 世纪 80 年代到 90 年代中期主导了手机功放市场，见图 11.7。注意到，除了栅源肖特基二极管以外，还存在着栅漏二极管。在正常工作时，两个二极管都是反向偏置的。

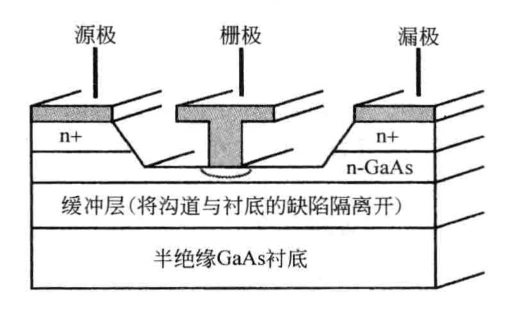


图 11.7　典型的 n 沟道 GaAs MESFET

对于现今的 MOSFET 和 MESFET 而言，每毫米栅宽的栅源电容大约在 1pF 的数量级，而固有的电压增益（$g_m r_{ds}$）一般在 10～20 之间。尽管会受到工艺影响，跨导的最大值一般为 1～2mS/mm，漏极电流的最大值大致为 1～2mA/mm。这些数字只是供数量级的估计。它们并不是自然常数，因此只能用于最为粗糙的手算。

如今，GaAs MESFET 在手机功放的应用基本上被 Shockley 的另一个预言，异质结双极型晶体管（HBT）所取代。普通的半导体器件对所有的层都采用同一种材料，这些层之间的仅有的电子学差别在于掺杂上有所不同。因此，所有这些经典器件都是同质结。大约在 1950 年，Shockley 就意识到将具有不同能带间隙的半导体材料结合起来会给器件设计带来功能强大的自由度[⑨]。以 BJT 为例，重掺杂的基区会减小基极寄生电阻，提高击穿电压和输出电阻。但这也会使电流增益下降。与此同时，在发射区重掺杂会提高电流增益却带来较大的发射结电容。如果在发射区使用宽带隙材料，这些矛盾能在相当程度上得以缓解。通过抑制基区到发射区的空穴注入，这样一个 HBT 能够在基区重掺杂、而发射区轻掺杂的情况下不至于使电流增益降低太多。由于减小了基区寄生电阻和发射结电容，HBT 的高频性能得到改善。基区的重掺杂同时减小了集电区-基区耗尽层向基区的延伸，这又减小了基区宽度调制效应（Early effect，厄利效应），从而增大了输出电阻。

虽然这个基本想法已经被提出数十年了，但直到 20 世纪 70 年代因为半导体技术的发展到足够先进的程度，才使得制造出这种器件成为可能。其中一个重要的限制是这些不同的半导体必须具有十分相同的晶格常数，否则在它们的接触面将充满缺陷。使用最为广泛的异质结双极晶体管是用 AlGaAs 制造发射区、GaAs 制造基区（集电区也是）[⑩]。锗硅（SiGe）HBT 同样也很受欢迎，尽管它不符合 Shockley 对 HBT 最初的定义。由于材料稳定性的关系，SiGe HBT 不能在基区中采用足够大的锗的摩尔分数，因而无法获得显著的带隙变化。然而，基区中的锗浓度梯度主要是用来增强已经存在的经典的因掺杂梯度而引起的

⑨ 美国专利号 # 2569347，1951 年 9 月 25 日授予。关于这个概念的最早公开发表的文章很明显是 Herbert Kroemer（"Theory of a Wide-Gap Emitter for Transistors"，*Proc. IRE*，November 1957，pp. 1535-1537）。他显然从文章的评审者那里得知了 Shockley 的专利。

⑩ 习惯上用构成衬底的材料来命名这种异质结技术。因而，用 AlGaAs 做发射区，而 GaAs 为衬底的 HBT 一般被称为 GaAs HBT。在那些场合下，那里有需要显式地注明发射区，我们也可以称为 GaAs/AlGaAs HBT。

电场。该电场对流过基区的载流子进行加速，从而提高了高频电流增益。SiGe 晶体管得以广泛应用是因为它与主流的集成电路制造工艺的兼容性，因为它与常规的双极型器件相差不大。

图 11.8 中的缓变层是用来缓解 GaAs 与 AlGaAs 界面上因小的晶格常数失配而引起的不十分严重的应力。缓变层通常会做得尽可能地薄，以减小它们可能引入的寄生串联电阻。

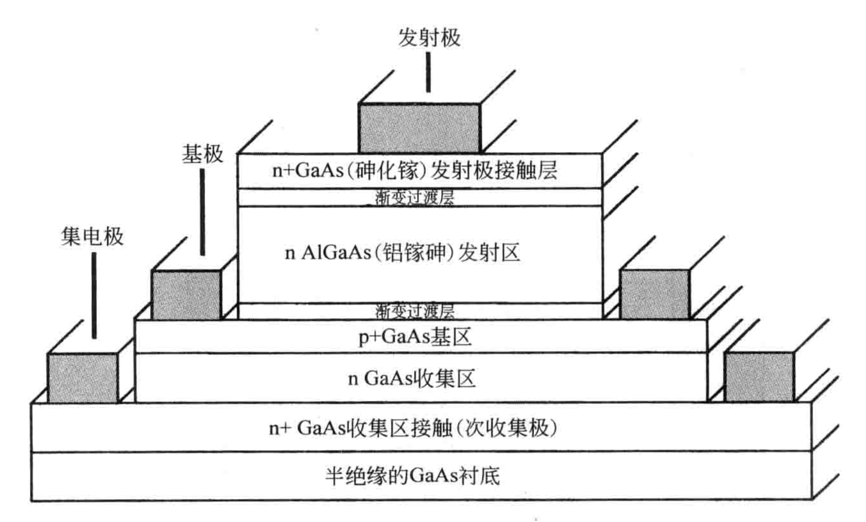


图 11.8　典型的 GaAs HBT(异质结双极型晶体管)(理想化情况)[11]

最近，磷化铟(InP)HBT 也已面世，这在一定程度上是因为该材料与红外射线的产生和检测的兼容性。很多用在射频应用中的 InP HBT 是双异质结型器件，它们的发射区是 InAIAs，基区是 lnGaAs，集电区是 InP。新引入的基-集异质结不仅提高了击穿电压，还减小了漏电流和输出电容。基区的 lnGaAs 与之前的 GaAs 相比，载流子迁移率提高了 50%。

相对于载流子横向流动的 MOSFET 和 MESFET，我们注意到 HBT 中载流子是在垂直方向流动的。当前的制造工艺(如金属有机气相沉积 MOCVD、分子束外延 MBE)无一例外采用图 11.8 所示的台面型结构。由于台面结构的一部分结是暴露在外的，这会缩短器件的工作寿命并带来高度可变的器件特性，硅的台面器件已被平面型晶体管所取代。然而截至目前，HBT 的制造工艺还没有出现类似的过渡。

通过引入异质结来调整器件的特性并不局限于 HBT 的实现。在 1980 年，高电子迁移率晶体管(HEMT)首次亮相，它是基于隧道二极管的发明者 Leo Esaki 在整个 20 世纪 70 年代所做的开创性的工作[12]。这种器件是建立在“载流子在无掺杂区域漂移最快”的内在了解之中的。不巧的是，一个普通的 MESFET 器件的阈值电压又是依赖于掺杂水平的。因此，在经典的 MESFET 中，阈值电压和迁移率高度关联。这使得人们不可能在不降低高频性能的情况下随意改变阈值电压。反之亦然，最大限度地提升高频性能可能会得到一个不方便的或根本不可用的阈值电压。

为了得到所需要的附加的自由度以解除以上两特性之间的耦合，可以使用适度掺杂的

---

[11] S. M. Sze(编辑)，*High-Speed Semiconductor Devices*(高速半导体器件)，Wiley, New York, 1990.

[12] 1973 年，Esaki 因发现重掺杂 p-n 结的隧道效应而获得诺贝尔物理学奖。

宽带隙材料(如 AlGaAs)，肖特基栅接触就是在宽带隙材料的下面使用了窄带隙的本征半导体(如 GaAs)构成。借助能带间隙的差别，载流子被很好地限制在这个本征层中(形成所谓的“二维电子气”)，同时通过调整宽带隙层的掺杂浓度来控制其他关键性的器件特性(如阈值电压)。由于载流子在未掺杂沟道中出色的传输性能，迁移率很容易被提高 1～2 倍，从而使得管子的跨导和高频性能得以提升类似的倍数。

位于非掺杂沟道上面的隔离层(见图 11.9)有助于将沟道与 n＋AlGaAs 层中的施主离子隔离开来。这一层的存在对保持沟道中载流子的高迁移率是必要的，但隔离层要做得尽可能薄以使进入沟道的电荷耦合最大。

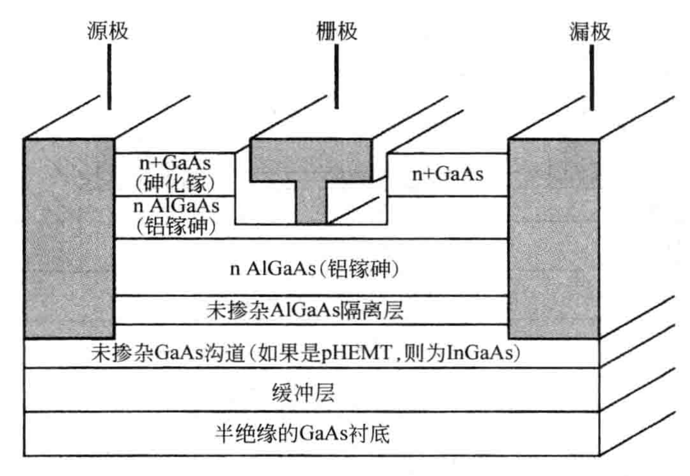


图 11.9　典型的 GaAs HEMT(高电子迁移率晶体管)[13]

HEMT 也有其他几个不同的称呼，如称为调制掺杂场效应晶体管(modulation-doped FET，MODFET)或异质结型场效应管(HFET)。其他不常用的缩写还有 TEGFET 或 2DEGFET(均表示二维电子气 FET)，以及 SDFET(选择性掺杂，selectively doped FET)。一项非正式的调查表明，HEMT 是最流行的叫法，而 HFET 和 MODFET 用的要少得多。

HEMT 又分成子类，包括赝配 HEMT(pseudomorphic HEMT，pHEMT)和渐变 HEMT(metamorphic HEMT，mHEMT)。赝配指的是利用有用的电学效应，这种电学效应是伴随因晶格常数的些微失配而引起的应力产生的(前提是该失配不至于产生严重的缺陷)。这些电学效应可能包括增大了的带隙差别和增强了的迁移率。在一个典型的 pHEMT 中，传统 HEMT 中的非掺杂高迁移率的 GaAs 沟道被迁移率更高的非掺杂 InGaAs 沟道取代，以获得更好的高频性能。基于 lnAlAs/InGaAs/lnP 这种材料组合，还可以有其他形式的变化。

渐变 HEMT(mHEMT)通过借助使用渐变层来容纳晶格失配的影响，因而扩展了可适用的材料组合的数目。这种能吸收应力的缓冲层允许在无掺杂的沟道中使用更大摩尔分数的铟。即使晶格失配产生了缺陷，它们也会仅限于缓冲层内，而这在大多数情况下不会对电学特性产生不良的影响。

由于 HEMT 是 MESFET 的修改版本，它们一般也是耗尽型器件。如果只是单电源供

⑬　见脚标 11。

电,对这种器件施加偏置是多少不方便的。事实上,正是这种加偏置的不便使手机上的功放转向了 HBT。由于不需要产生负电压的电路,HBT 可以让手机做的更小,成本更低,电池寿命更长。

另一个值得注意的性质是 GaAs 和 InP 的热传导率远比硅差。硅的热传导率是 GaAs 的三倍,InP 的两倍。这种差异在功放中显得尤为重要。即使在所谓的小功率系统中,有时瞬时功率也可能会大到足以使热效应表现出来。那些习惯于只与硅器件打交道的工程师经常会对设计中必须考虑到的热效应的重要性感到吃惊。

最后,不得不说,器件工程师丰富的创造力绝不会枯竭。因此,我们在前文中提供的不完整器件类型单子只会变得更不完整。比如,现在对碳化硅(SiC)和氮化镓(GaN)器件的研究就非常活跃,我们可以预期,未来采用的材料将远超这里提及的种类。

11.2　建模

有源器件一般都表现出很强的非线性,而且它们的许多器件参数很容易受到频率、温度和偏置的影响而变化。在电路分析中包括非线性已经是非常具有挑战性了,又不得不在分析中考虑电抗就更加难上加难。尽管有廉价的辅助计算可包含器件模型中的“所有参数”,但这仅仅是有利于电路分析,而不是电路设计。

另一种方法在很大程度上不需要理解内在的物理知识,设计主要依赖于测量得到的参数(如 S 参数)。因为它与实际相结合,这种方法能得到不错的结果,因为是基于“真实”的数据。但是由于缺乏基于物理的紧凑模型,为了进行鲁棒的设计需要大量的数据组。在数据组之外(即使是属于同一类晶体管中的不同器件)要对管子的性能做出预测也是不太可能的。

我们经常会听到抱怨说这个或那个模型是“错”的。能认识到所有的模型在某种程度上都会有错误,这一点很重要,正因为如此,寻求正确模型的努力,才永不会停止。从另一个角度说,正确的工程理念是去找到那些能回答需要答案的模型。如果问题本身是简单的,那么简单的模型就足够了。超过实际工作所需求的过度复杂的模型,只会让人们更难把握问题的本质。

在为设计而建模时的一个永恒主题是要通过有选择地、下意识地忽略那些被认定为二阶效应的现象,来降低模型的复杂度。当然,这中间的窍门是认定那些对于特定情形的、真正的二阶效应。同样,认识模型有效的范围也非常重要。这样就能在模型能正确工作的范围内使用它们。

本章的目的是回顾一些基于物理而又相对简单的模型。这样的模型适用于电路的初步设计。当更为全面的仿真或者实验测量表明电路需要修改时,这些模型对设计的进一步改进是不可或缺的。没有这些简单模型的引导,设计者将很难判定哪些参数需要调整,是该调大还是调小,以及调整量是多少。

我们将给出 FET 和 BJT 的小信号模型。对于 BJT,我们将假设集电区-基区之间的结不会正向偏置(即晶体管不会饱和)。此外,我们还会花一点时间来讨论如何从看起来没有头绪的数据表中提取关键的模型参数。

11.3 双极型晶体管的小信号模型

11.3.1 简单的直流模型

对于工作频率足够低以至于晶体管内的电抗成分可以忽略，同时对在集电极-发射极之间施加的电压不变的情况下，晶体管集电极电流(i_C)与基极-发射极电压(v_{BE})之间的关系可以用一个指数函数来很好地近似：

$$i_C \approx I_S\left[\exp\left(\frac{v_{BE}}{V_T}\right)\right] \tag{1}$$

其中 V_T 表示热电压，它等于 kT/q，在室温下约等于 25mV。令人惊讶的是，这个指数关系是如此之准确，对实际的硅器件在集电极电流变化 6～8 个数量级范围内，误差只有几个百分比。从式(1)可以看出 v_{BE} 每增加 18mV，i_C 增大一倍；而 v_{BE} 增加 60mV，i_C 变成原来的 10 倍，记住这些关系有时会有用。

参数 I_S 表示饱和电流，在有些教材中它的写法容易使读者误以为它是一个常量。事实上，它的行为要复杂些：

$$I_S = I_0\exp\left(-\frac{V_{G0}}{V_T}\right) \tag{2}$$

其中 V_{G0} 是带隙电压(外推到绝对温度 0K 的值。就硅而言，这个值约为 1.2V)。I_0 的一个准经验公式为

$$I_0 = A_e BT^r \tag{3}$$

其中 A_e 表示发射区的面积，B 是个与工艺相关的常量，T 仍表示绝对温度，r 也是个工艺相关的量，我们称为曲率系数。较早的双极型工艺发射区是深度扩散的，r 的典型值在 2～3 之间；而现代的高速半导体工艺常采用浅的用离子注入形成的(重掺杂[14])扩散区，r 的典型值接近 4。显然，I_S 是与温度和工艺相关的“常量”。

在式(1)中假定晶体管工作在正向有源工作区，此时发射结正偏，集电结反偏。还假定集电极电流远大于 I_S，这些假设对所有工作在正常模式下的微波器件均成立。

从式(1)～(3)可以看出，i_C 不变时，温度每升高一度，v_{BE} 就下降 2mV，如图 11.10 所示。单独应用式(1)会得到一个错误的温度系数符号，这也强调了使用 I_S 的详细表达式的重要性。一个在温度行为上的小曲率并没有在图中显示出来。事实上这个曲率是非常小的，以至于除了在特定的电压基准源(所谓的带隙基准源)情况外，其存在可以被忽略。

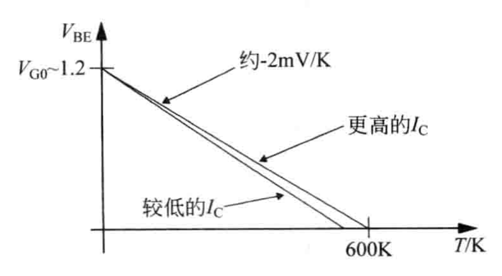


图 11.10 集电极电流为常数时，v_{BE} 随温度的近似变化关系

[14] 这些高的 r 值可能是因重掺杂发射区的能带隙变窄与非线性性引起。

集电极电流与基极电流之间的关系也很重要：

$$I_C = \beta_F I_B \tag{4}$$

在器件设计时为了获得好的高频特性而做出的折中会导致晶体管的 β 值低于用于低频的晶体管的值。β 的标称值通常在 100 以下，甚至会低至 10。要记住，不同管子的 β 值可能会相差很大，即使同一个管子的 β 值也会随电流和温度呈现较大变化。

显然，i_C 与 v_{BE} 之间的关系是非线性的。如果电路中包含一个非线性关系，分析起来就会很复杂了；如果要分析多个非线性电路，人们就会感到更无助了。因此，线性化模型（又称增量模型、小信号模型）被发展起来来简化分析，尽管这种模型只可用于有限的范围内。

在 BJT 的情况下，建立一个合适模型的第一步是把发射结电压看作是一个直流分量上叠加一个小信号分量：

$$v_{BE} = V_{BE} + v_{be} \tag{5}$$

将式(5)代入式(1)，然后进行泰勒(Taylor)级数展开得到

$$i_C \approx I_S\left[\exp\left(\frac{V_{BE}+v_{be}}{V_T}\right)\right] = I_S\left[\exp\left(\frac{V_{BE}}{V_T}\right)\right]\left[1+\left(\frac{v_{be}}{V_T}\right)+\frac{1}{2!}\left(\frac{v_{be}}{V_T}\right)^2+\cdots\right] \tag{6}$$

如果忽略所有的非线性项（记住，对工程师而言，效率第一），就得到了一个线性关系：

$$i_C \approx I_C\left[1+\frac{v_{be}}{V_T}\right] = I_C + \left(\frac{I_C}{V_T}\right)v_{be} = I_C + i_C \tag{7}$$

这说明总的集电极电流包含一个直流项和一个与小信号发射结电压大约成正比的项：

$$i_C = \left(\frac{I_C}{V_T}\right)v_{be} = g_m v_{be} \tag{8}$$

比例常数 g_m 称为跨导（“跨”是因为电流和电压分别在不同的端口测得；“导”即电导，因为它是电流除以电压），因此跨导用一个压控电流源来进行建模。

集电极和基极电流之间也有关系。这样，我们就得到基极电流与基极-发射极电压之间的关系：

$$i_b = \frac{i_C}{\beta_0} = \left(\frac{g_m}{\beta_0}\right)v_{be} = \frac{v_{be}}{r_\pi} \tag{9}$$

注意上式中使用 β_0 是为了强调集电极-基极的小信号电流比可能会与直流下的值不同。但多数情况下，我们认为 β_0 和 β_F 是相等的。通常情况下，这种替代是合乎常理的，这是因为 β 非常易变，过分挑剔它的值是没有意义的。

基极电流和发射结电压之间的小信号线性关系可以简单地用一个在基极和发射极之间的电阻(r_π)来表征。

在所有实际的双极型晶体管中，集电极电流并不只是基极与发射极之间电压的函数，它也会受集电极与发射极之间电压的弱的影响。这是因为集电极-基极耗尽区的宽度随着集电极电压的增大而增加，因此基区的有效电宽度同样依赖于集电极电压。由于集电极电流与基区宽度相关，集电极电压的增加会带来集电极电流的增大，这一现象称为基区宽度调制效应（也叫厄利效应，因为贝尔实验室的工程师 James Early（厄利）首先发现并描述了这一现象）。这个现象用集电极-发射极电阻 r_0 来建模[15]。如果将一组集电极电流作为集电极-

[15] 基极电流也会减小，这可以通过在集电极和基极间增加一个电阻来建模。在微波电路的典型工作条件下，这个电阻大多数情况下都可以忽略。如果对更精确的模型感兴趣的话，可以参考 Ian Getreu 写的经典书籍：Modeling the Bipolar Transistor (Tectronix, Beaverton, OR, 1979)，或者 P. Gray 和 R. Meyer 合著的 The Design and Analysis of Analog Integrated Circuits (Wiley, New York, 1996)。

发射极电压的函数绘在同一张图上（V_{BE}为参量，译者注），则曲线的向上倾斜就反映了基区宽度调制效应。把所有这些曲线反向延伸时，它们在V_{CE}轴上近似地相交于同一点，这个截距（其负数）被称为（定义为）厄利电压V_A。对于给定的集电极偏置电流，集电极-发射极电阻值：

$$r_o = V_A / I_C \tag{10}$$

绝大多数微波双极型晶体管的厄利电压典型值在10～50V之间。

为了完成模型的直流部分，注意到晶体管是由电阻率不为零的半导体材料制成的，因此在每个端子上存在一些串联电阻。多数情况下，基极电阻最为重要，所以我们通常忽略另外两个端子上的串联电阻（然而，当这些电阻起一定作用时，这种忽略的有效性需要进一步验证）。综合以上分析，我们得到BJT的小信号模型，如图11.11所示。

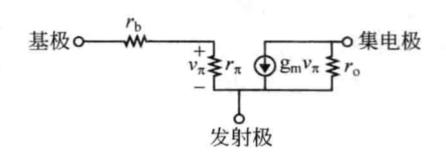


图11.11 双极型晶体管的低频混合π模型

与命名有关的一些知识："π"源于模型的形状（当然，这需要一点儿想象力），而"混合"指的是模型中的参数并没有使用同一量纲，此处，我们用到了电阻和（跨）电导。

就像我们以后在考虑FET时会看到的那样，所有这些晶体管模型有着相同的基本拓扑结构。理由很简单：目前我们使用的所有晶体管本质上都是压控电流源。也就是说，它们都是跨导器[16]，而小信号模型刚好能反映这一事实。各模型的区别主要是连接在跨导器各端子上的阻抗性质不同。

现在我们来看一下，模型的"小信号"部分是怎么样的？换言之，信号要多小才算"小信号"？当理解了对泰勒展开式尾项的截断是整个线性化过程的核心，答案就不言自明了。因此，当被忽略项相对于这个主要的线性项的比到达一定的百分比时，可以认为信号就不再算"小"了。任取一个具体数值，假设当二阶项不超过一阶项的10%时可以满足（精度）要求。在这种情形下，要求

$$\frac{1}{2!}\left(\frac{v_{be}}{V_T}\right)^2 < \frac{1}{10}\left(\frac{v_{be}}{V_T}\right) \Rightarrow \left(\frac{v_{be}}{V_T}\right) < \frac{1}{5} \tag{11}$$

因此，室温下当发射极-基极电压变化超过5mV时，小信号分析就开始变得不可靠了。注意到"小"与"大"（信号）的界限比较模糊，因为所容许的误差本身是待解决问题的函数。即便如此，对双极型晶体管而言，"小"信号也绝不会是好多个毫伏值。

11.3.2 一个简单的高频模型

为了建立一个简单的模型，考虑一个更能代表实际器件的截面图是非常有用的，如图11.12所示。尽管此截面图本身已经是简化了的，但与之前的NPN三明治模型相比还要详细得多。这里电流总体上是纵向流动，虽然在载流子向上流到集电极之前存在着横向流的分量。通常，为了降低相应的集电极串联寄生电阻，图11.12中在发射极的右侧再放置

---

[16] 许多教材将BJT描述成电流控制器件，这一观点尽管能在数学上得到支撑，但是它偏离了一个真实的物理：从根本上来说，是发射结电压降低了发射结的势垒高度，向基区注入载流子，这些载流子从基区扩散（漂移）到集电极。基极电流的理想值是零，所以在基极呈现的电流都是寄生电流。因此电流控制电流源的说法很难被认为是对基本控制变量的正确描述。

一个额外的集电极接触。通过对高阻的外延集电区平行地放置一个低阻支路，这个重掺杂埋层可以减小这个电阻。在发射极接触的右侧添加另一个基极接触也能够减小基区寄生电阻。

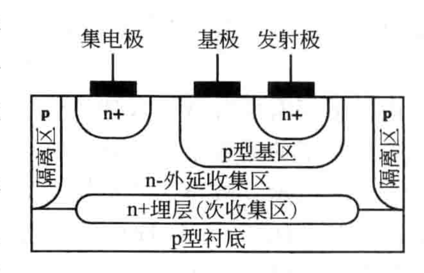


图 11.12 更为典型的集成电路中的双极型晶体管结构（各层未按比例绘制）

在高频时，至少需要考虑那些依赖于频率的器件行为。为此，首先注意到双极型晶体管是由两个结构成。发射结是正向偏置的，而集电结是反向偏置的。因此双极型晶体管至少存在两个结电容（分别位于基极-发射极对和集电极-基极对）。

小信号电容可以很好地近似为

$$C_j = \frac{C_{j0}}{(1 - V_j/\psi_0)^m} \tag{12}$$

其中 C_{j0} 是零偏置时的结电容，V_j 是结电压（正向偏置时为正），ψ_0 是标称的接触电势（即自建电势，译者注）（取决于材料的能带间隙，硅材料的典型值在 1V 左右），m 是一个与掺杂分布相关的常量。对于线性渐变结，m 是 1/3；突变结 m 为 1/2。采用超突变结能获得大于 1/2 的 m 值（在第 9 章中我们曾详细讨论过）。

在反向偏置以及较弱的正向偏置下，结电容的值非常符合式(12)。在深度正偏时，结电容的值通常会被限制在 2～3 倍的 C_{j0}，而不是随着结电压趋于 ψ_0 趋向无穷大。在放大器中，有必要考虑结电容与电压的相关性，因为这会带来失真。有些特殊情况下，我们也会利用这种非线性，比如用它来产生谐波。

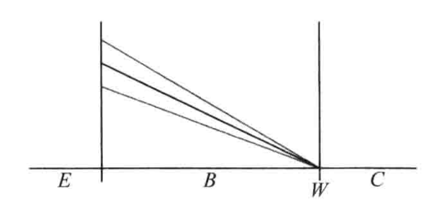


图 11.13 不同基极-发射极电压下的基区过剩载流子浓度

在反向偏置的集电结中，平板电容一直存在。然而对于正向偏置的发射结，我们还有更多的工作要做。为了理解为何除平板电容（由式(12)给出）之外，在这个条件下还存在其他电容，参见图 11.13：图中简单（而又十分近似地）描绘了过剩载流子浓度在基区中位置的函数。

在与发射区相接的边界处的注入载流子浓度与总的发射结电压是成指数关系的。图 11.13 中所示的是 v_{BE} 取 3 个不同值时过剩载流子浓度的带有假定性分布描绘。忽略载流子的复合将得到理想的直线分布。由于集电结的反向偏置，在与集电区相接的边缘载流子浓度变为零。载流子从高浓度区（位于发射区边缘）主要通过扩散流向集电区边缘（此处载流子浓度为零[17]）。

在图 11.13 中，对任意一个给定的 v_{BE}，基区的总电荷量正比于三角形的面积。当 v_{BE} 增加时，额外的电荷需要通过基极被提供；当 v_{BE} 减小时，电荷需要被外界抽取。这种特性可以通过电容来建立模型。但不同于普通的结电容，该电容的大小正比于 v_{BE}（对于固定的 v_{BE} 变化量，v_{BE} 越大，则电荷三角形的变化量就越大）。考虑线性化时，这种比例关系意味着一个电容，我们称之为扩散电容 C_b，正比于集电极的直流偏置电流。因此，发射结电容可以表

⑰ 在几乎所有的现代晶体管中，基区掺杂浓度都是随位置变化而变化的，从而产生了内建电场，能大大加快载流子穿过基区。该电场会阻碍反方向的载流子漂移，这也解释了为什么 BJT 在集电极和发射极互换时高频性能大大恶化。

示成常见的结电容（与电压相关）与该扩散电容（电流相关）之和：

$$C_\pi = C_{je} + C_b \tag{13}$$

考虑以上所有电容就能得到如图 11.14 所示的模型[18]。

封装了的器件带有额外的寄生参数，其中影响最大的是串联在各个端子上的电感，如图 11.15 所示。减小这些不需要的寄生电感是一个挑战，额外的发射极寄生电感特别不受欢迎。举个例子，一个 1nH 的发射极电感在 1GHz 时会呈现超过 6Ω 的阻抗。这个固有的阻抗会显著降低该级电路的增益，并且很可能会使电路发生谐振，从而破坏增益平坦度。许多高频封装会提供至少两条的发射极引线，这正是为了减小有效串联电感。

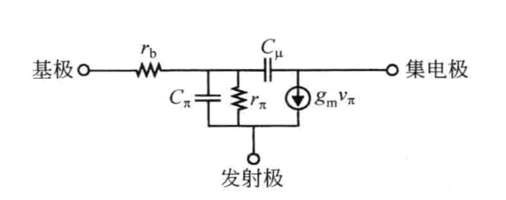


图 11.14　双极型晶体管的更高频混合 π 模型

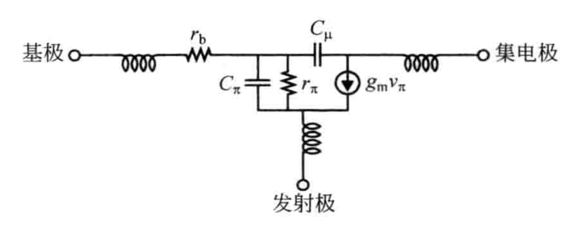


图 11.15　包括某些寄生元件的 BJT 的高频混合 π 模型

更为详尽的有关封装带来的寄生模型还会包括各种电阻（表征引线损耗）和电容（各个终端之间）。然而，在图 11.14 中已经考虑了射频微波工程师最为关心的大多数现象。再次强调一下，我们寻求的是能为我们提供设计的内在见解的模型，而不一定要有正确的结果。采用了更为精确的模型仿真器会帮我们解决这些问题。

11.3.3　高频品质因数

一个常用的品质因数是当输出交流短路时共发射极电路的电流增益外推到 1 时的频率。对图 11.16 中的电路而言[19]，

$$\omega_T = \frac{g_m}{C_\pi + C_\mu} \tag{14}$$

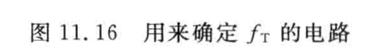


图 11.16　用来确定 f_T 的电路

注意到这个品质因数不受基极串联电阻的影响。因为基极是由电流源驱动的[20]，这使得串联在基极上的任何阻抗都与 ω_T 无关。然而，实际的电路性能肯定会受这些阻抗影响，这揭示了这个特别的（且被使用过度的）测量方法的缺陷。经常会发生这样的情况，即以牺牲在实际中更为有用的参数为代价来最大化这个参数。另一个问题是任何与集电极-发射极两端并联的阻抗是同样地被忽略了，这是因为在测量时，这两个端是被交流短路的。

由于电流增益并不具有根本性质上的含义（比如说，原则上总是可以借助理想变压器在不消耗能量的情况下获得任意大的电流增益），而且考虑到 ω_T 在其他方面是个不完善的品

⑱　注意到此处忽略了输出电阻 r_0。尽管这样做带来的误差通常可以忽略，但是在任何输出电阻可能会带来影响的情况下，需要检查该假设的有效性。这些例子包括高增益电路和低噪声放大器，因为它们的输出阻抗容易受 r_0 影响。

⑲　要直接从这个电路推导出这个公式，C_μ 对输出电流的贡献被忽略，但其对输入电流的贡献并没有被忽略。

⑳　“电流源串联一辆别克（汽车品牌）汽车仍是个电流源。”

质因数，我们引入一个不同的衡量器件性能的标准来对 ω_T 加以补充。这就是最大振荡频率 ω_{max}，用来表征晶体管的功率增益下降到 1（或者外推到 1）时的频率。对我们的简化模型，有

$$\omega_{max} = \frac{1}{2}\sqrt{\omega_T / r_b C_\mu} \tag{15}$$

这个衡量标准更令人满意，因为它显式地反映了基极串联电阻对晶体管高频性能的影响。

为了测量 ω_T 和 ω_{max}，需要绘出不同频率下的各种有关的增益，直到曲线呈现明显的单极点走向[21]。每条曲线向外延伸就得到单位增益频率。这种外推法是必要的，因为随着频率升高，晶体管的行为会更加复杂。就 ω_T 而言，由于 C_μ 馈通的影响，采用我们的模型可能永远也无法获得单位增益。

尽管前面的讨论均是基于硅基双极型晶体管，基本模型亦适用于异质结双极型晶体管（HBT），至少就这些模型提出的初衷而言（即获取设计的内在见解）。当然，模型中的元件参数会各不相同。

11.3.4 从数据手册中提取模型参数

尽管（后者说是因为）在数据手册中包含了大量的数据，提取模型参数是一件很具挑战性的事。我们再一次提醒，在建模过程中所说的与做的，在某种程度上都是带有错误的，但我们依然能提出一种计算简单且足够准确的参数提取方法。我们感兴趣的模型参数是电容 C_μ、C_π 和基极的寄生电阻 r_b。我们也许还会对串联在各个端子上的电感感兴趣。有时，我们也会关心各端子间的互感（比如在基站中的射频功率管，它们一般是由多个器件并联构成）。

数据手册中通常会直接给出 C_μ 的值，但通常标为 C_{ob}，这样命名是因为在测量它时采用了发射极开路的共基极测试电路。快速回顾一下我们的晶体管模型就会发现，这种测试中实际测得的电容正是我们所说的 C_μ。

通常也会给出 C_π 和 C_{je} 中的结电容部分的值。只不过它们几乎都是在反向偏置下测得的。鉴于发射结工作在正向偏置下，数据手册中的 C_{je} 的值必须通过某种方式换算为正向偏置下的值。最简单的方法就是找到 C_{je} 在零偏时的值，然后乘以 2 即可。（式(12)不能作为换算的依据，因为它在通常的正向偏置下不成立。）幸运的是，C_{je} 很少起主导作用，因此即使存在较大的误差也可以接受。

扩散电容项 C_b，最好是从 ω_T 和集电极电流的函数关系中得出（见式(14)）。作为一个粗略的近似，可以通过数据手册提供的参数来提取单位电流的扩散电容估计值，为

$$\frac{C_b}{I_c} \approx \frac{1}{V_T \omega_{T,pk}} \tag{16}$$

式中 $\omega_{T,pk}$ 表示 ω_T 的最大值，V_T 是热电压（kT/q），I_c 则是获得 $\omega_{T,pk}$ 时对应的集电极电流。

在低电流工作模式下，扩散电容可以忽略不计，此时总电容保持不变，ω_T 就随 g_m 线性增加，因而 ω_T 与集电极电流有同样关系。最终，扩散电容将决定分母大小，导致 ω_T 变为一个常量。此时在 ω_T 为常量时，C_b 可以合理近似为 g_m/ω_T。在实际的器件中，由于大注入效

---

[21] 电流增益按 $1/f$ 下降，而功率增益按 $1/f^2$ 下降。

应使得基区变宽以致电荷量以超线性关系增加，ω_T 最终会随着 I_c 的增加而降低，因此扩散电容的确定要选在截止频率与集电极电流关系最平坦的区内。在实际的 HBT 中，大注入效应受到抑制，以至于对平坦区域的选择比对一般同质结双极晶体管来得简单。

推断 r_b 是最困难的任务，并且要接受得到值有相当大的不确定性这一事实。在第 13 章会讲到，输入阻抗的实部包括 r_b 和由发射极电感和晶体管之间相互作用产生的感生电阻：

$$r_{in} \approx r_b + \omega_T L_e \tag{17}$$

大多数微波晶体管的数据手册都会给出输入阻抗与频率之间的函数关系曲线。首先找到输入谐振点，记下此时的电阻为 r_{in}。运用这个谐振频率和已知的器件电容推导出总的封装电感。由于总的封装电感是基极和发射极电感之和，我们需要通过判断或者计算如何将总电感划分到两个端子上。对许多封装来说，对半分是不错的选择。最后，借助已知的 ω_T 来完成 r_b 的最终计算。遗憾的是，$\omega_T L_e$ 往往是 r_{in} 的主要成分，以致 r_b 的估算存在极大不确定性。

注意到 r_b 会随着偏置点而变化也是非常有帮助的。集电极电流较低时，相应的基极电流也较低。因此，在基极串联电阻的压降较小，整个发射结界面在将载流子注入到基区方面是有效的。在大电流时，基极电阻上的压降变大。由于这个压降，横跨发射结上的电压沿发射结界面变化：从表面到深处这个电压降逐渐变小，这是由于基极和发射极接触之间的距离在结的表面处较短。从而，发射结界面注入了不均匀的电流，越接近表面电流越大。基极接触和有效的发射区之间的平均距离因而减小，导致了 r_b 减小。r_b 最大值与最小值之比的典型值是 2。

11.4 场效应晶体管(FET)模型

幸运的是，MOSFET、MESFET 和 HEMT 可共用同样的模型(至零阶的精度)。如前所述，它们的模型与双极型晶体管的十分相似，如图 11.17 所示。我们有意地避免给出任何详细的大信号 V-I 方程，这是因为，对于现代 MOSFET 和 MESFET 来说，二阶效应大到足以使任何一组简单的方程定量说来无用。甚至从定性的角度来讲，所谓“电流随着栅源电压的二次方变化”的说法也是不正确的。因此，我们会给出一些定性和准定量的观察。

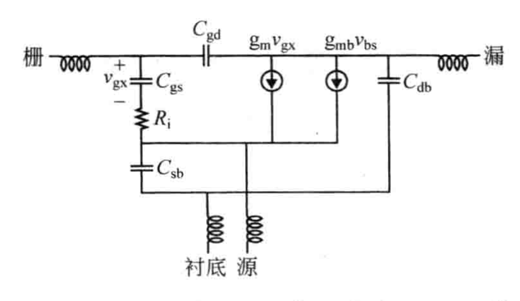


图 11.17 简单的增量(即小信号，译者注)FET 模型

首先，所有的 FET 都通过调制半导体的导电率来工作。对于耗尽型器件，如 JFET 或者典型的 MESFET/HEMT，默认状态下器件是导电的，因此，给器件加上一个控制电压来

减小其导电性。对于 n 沟道器件，栅源电压是负的[22]。我们所知道的一个参数“夹断”电压对应于产生外推漏极电流为零时的栅源电压。在增强型器件中，默认的状态是非导通的，因此必须加一个有效的栅源电压以使导通。一旦栅源电压超过某个特定的阈值电压后晶体管就开始导通了。从这些描述中可以看到，阈值电压和夹断电压是描述相同的事情，只是分别从相反的默认条件得到的。

FET 有两种不同的工作机制。工作于三极管区域时，与它的始祖真空管一样，FET 表现为一个电压控制的(尽管是非线性的)电阻。栅源电压较小时，这个电阻反比于栅极的过驱动电压(栅源电压超过阈值电压的那部分电压)：

$$r_{ds} \propto 1/(v_{GS} - v_T) \tag{18}$$

(注意，这里 v_T 表示阈值电压，不是热电压，译者注)只要漏源电压小于栅源电压，FET 就会工作在这个电阻区。对于更高的漏源电压，栅极电压在靠近器件漏极的沟道部分不足以感生出载流子。导电沟道就到达了所谓的夹断状态。即便如此，仍然有电流流过。可能这样的陈述会使你困惑，因为它基于非物理的巧妙处理方法。电荷密度降到零，但这些电荷流动的速度无限地快(现在明白了吧?)。由于“无限接近于零”的不可思议的特性，其乘积是有限的(或者等于任何你想要的值)。实际上，真正的夹断不会出现，电荷密度在(沟道的，译者注)漏极末端开始降低，但却不会等于零。同时，载流子速度也不可能达到无限，只会达到由材料特性决定的最大速率：饱和速率。对于给定的材料[23]，这是载流子能够达到的最高速率。砷化镓的饱和速率大约是硅的两倍，这就是为什么砷化镓广泛应用于 RF 和微波电路的原因。

一旦夹断(这是在一定程度上的说法)发生，漏极电流就会饱和(因为速率和电荷密度都会饱和)。因而就很自然地把这个工作区域称为饱和区。在那些一直说不太清楚的经典的教科书式的 FET 里，工作在饱和区的漏极电流以平方率关系依赖于栅极过驱动电压：

$$i_D = K(v_{GS} - v_T)^2 \tag{19}$$

其中，K 取什么值是由保证漏极电流的表达式正确而决定的。对于 JFET，传统上 K 的表达式为

$$K = I_{DSS}/V_P^2 \tag{20}$$

其中，I_{DSS}是栅源电压为零时的漏极电流。在这个情形下，公式(19) 中的阈值电压用夹断电压来代替。

而对于双极型晶体管，这个电流并非真正地呈现理想的饱和，存在着对漏源电压的依赖。这个效应同样可以采用在漏源之间增加一个电阻 r_o 的方法来表征。

在此后的讨论中，我们集中在小信号的图像，而把大信号细节留给更加复杂的仿真器模型。或许注意到以下这一点就够了，即 MESFET 器件里有栅源和栅漏两个肖特基二极管，在器件的正常工作状态下，它们都处于反向偏置。然而，在非正常工作状态下，任一或者两

[22] 几乎所有的微波器件都设计成电子承载电流，因此，微波双极性晶体管通常是 NPN 器件，而微波 FET 通常是 N 沟道晶体管。原因就是电子的迁移率总是比空穴的迁移率大。在硅材料中，其不一致的比率是 2∶1，这使得 PNP 或者 P 沟道 FET 还可以被使用。然而对于其他材料，其不一致性要超过一个数量级，因此，比如 P 沟道商用的 GaAs MESFET 是不存在的。

[23] 这种说法只有在电荷沿着一条至少几个平均自由程(碰撞或者散射之间的平均距离)长的路径运动时才正确。如果路径足够短，载流子就会弹道式地运动，那么饱和速度的概念就不成立了。

个二极管都会被正向偏置。相比而言，MOSFET 没有与栅极直接相连的结。

注意在我们的小信号模型里有两个跨导。一个典型的 MOSFET 器件实际上有两个栅极。一个是常规的栅极，另外一个是衬底。因为这两个端子上电压的变化都可以调制漏极电流，因此在模型里面就需要两个独立跨导。如果源极端和衬底端连在一起（正如所有分立器件一样），这个模型将会被简化，这是因为背栅效应及其相应的模型元件都不存在了。

跨导可以用很多种方法来表示，但简洁的表达式如下：

$$g_m = B\frac{I_D}{V_{OD}} \tag{21}$$

这里，I_D 是偏置漏电流，V_{OD}是栅过驱动电压。当 V_{OD}大于几个 kT/q 时，这个公式有效，因此，就每单位电流对应的跨导而言，FET 的（不管是哪一种）跨导一定不如双极型晶体管（对 BJT，(21)式的分母就是 kT/q，译者注）。对于平方律关系器件，参数 B 的理想值是 2，实际器件中 B 则接近于 1。

图 11.17 所示的简单模型没有把端口的串联电阻表示出来。因为在 MOSFET 器件中，栅电极一般都用薄多晶硅制作，其串联栅极电阻通常不能忽略。器件设计工程师花很大的气力把硅 MOSFET 细分成几个并联的部分，就是为了使得总的栅极电阻比较低。由于 MESFET 器件的栅极采用相对较低电阻率的金属制作，其串联栅电阻也相应地比 MOSFET 的低很多。

对于 MESFET（包括这一类的其他器件）的情况，栅极电容的充电电阻 R_i 有时显得很重要。主跨导取决于栅源电容上所加的电压，因而 R_i 的存在就产生了一个低通滤波器。R_i 的典型值范围是 5～10Ω。

11.4.1 动态元件

电容限制了 FET 电路的高频性能，正如它限制双极型晶体管电路一样，所以我们必须知道这些电容的来源，粗略地估计它们的大小。

首先，由于源区和漏区与衬底形成了一个反偏结，我们可以想到这两个区域中的任何一个与衬底间都会有一个标准的结电容。这些电容表示为 C_{jsb}和 C_{jdb}，如图 11.18 所示，图中耗尽区的范围被放大了很多。

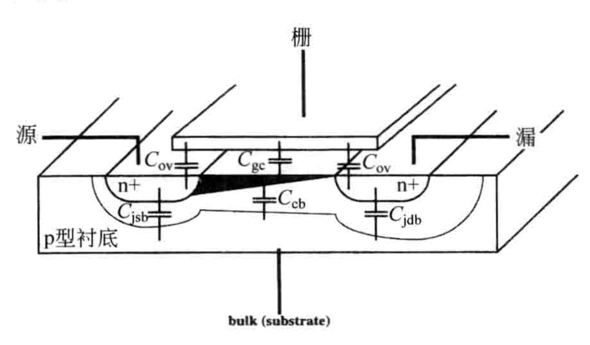


图 11.18　MOSFET 中的各种电容

除了结电容外，还存在其他各种平行板电容。在图 11.8 中的电容 C_{ov}代表着栅-源和栅-漏间的交叠电容，这些电容是我们非常不希望有的，但却是不可避免的。在制造工艺中，

源区和漏区的横向扩散长度与它们的纵向扩散深度很接近。因此,在工艺过程中,源漏扩散区会向外扩展并向栅极下延伸一定程度,作为一个粗略的估计,我们可以认为交叠量 L_D 为源-漏扩散深度的 2/3～3/4。因此,

$$C_{ov} \approx \frac{\varepsilon_{ox}}{t_{ox}} W L_D = 0.7 C_{ox} W x_j \tag{22}$$

式中 x_j 是源-漏区扩散的深度,ε_{ox} 是氧化层的介电常数(大约 $3.9\varepsilon_0$),而 t_{ox} 是氧化层的厚度。

由于平行板交叠电容随着边缘效应的增大而增大,因此即使不存在物理上的交叠情况,"交叠"电容这一项也不会为零。就这一意义上讲,我们应当记住,在现代器件中,栅极厚度实际上比沟道长度还大,所以图 11.18 所画的相对尺寸一种误导。要把实际的栅电极想象成是一个大的橡树而不是一块薄板。另外,源端和漏端的互连线尺寸不能被认为是可以忽略的。事实上,由于栅电极的厚度几乎没有随工艺进步而缩小多少(如果不是完全没有缩小的话),所以现在从一代工艺到下一代工艺,"交叠"电容的变化是十分缓慢的。

另一个平行板电容是栅-沟道间的电容 C_{gc}。由于源区和漏区都延伸到了栅下面的区域,所以有效沟道长度减少了两倍的横向扩散距离 L_D。因此 C_{gc} 的总值为

$$C_{gc} = C_{ox} W (L - 2L_D) \tag{23}$$

在强反型层,表面的电荷载流子和衬底的载流子类型相反。它们之间存在一个耗尽区。因此,沟道与衬底之间还存在一个电容 C_{cb},它的特性也像是一个结电容。它的值近似为

$$C_{cb} \approx \frac{\varepsilon_{Si}}{x_d} W (L - 2L_D) \tag{24}$$

其中,x_d 是耗尽层的深度,它的值为

$$x_d = \sqrt{\frac{2\varepsilon_{Si}}{qN_{sub}} \mid \phi_s - \phi_F \mid} \tag{25}$$

式中在绝对值符号中的量是衬底中的表面势和费米势之差(这句话的理解要特别注意,下面有关强反型的叙述是正确的,译者注)。在强反型区(包括三极管区和饱和区两者),这一量的大小为费米势的两倍。

然而沟道不是器件的引出端,所以如果要发现以上各项电容如何影响各引出端电容的话,就要牵涉沟道电荷是如何在源和漏之间划分的问题。一般来说,引出端的电容值取决于工作状态,因为偏置条件会影响该电荷划分。例如,当没有任何反型电荷(器件截止)时,栅-源和栅-漏电容就可以合理地近似为只有"交叠"电容这一项。

在线性区,器件中存在反型层,我们可以假设源和漏同等地分享沟道电荷。因此 C_{gc} 的一半计入到交叠电容这一项中。类似地,在线性区结电容 C_{jsb} 和 C_{jdb} 这两项都分别加上了 $C_{cb}/2$。

在饱和区,漏区的电势变化并不会影响沟道电荷。因此,C_{gc} 对 C_{gd} 的构成没有影响;C_{gd} 全部由交叠电容构成。栅-源电容是受到 C_{gc} 的影响,但"详细分析"[24]表明,大约只有 2/3 的 C_{gc} 加入到交叠电容这一项中。同样,在饱和区时,C_{cb} 对 C_{db} 也没有任何影响,但 C_{sb} 而言有

---

[24] 因子 2/3 来源于沟道电荷的计算,从根本上说是来自在平方率工作区下对图 11.18 中所假定的三角形分布进行积分的结果。

C_{cb}的 2/3 加入到其中。

栅-衬底电容在强反型层都可以认为是零(这对线性区和饱和区都一样,沟道电荷实质上起到了屏蔽作用,使得衬底免受栅极上电压变化的影响)。然而当器件处于截止状态时,存在着一个与栅电压有关的电容,它的值大致以线性方式在C_{gc}以及C_{gc}与C_{cb}的串联值之间变化。在低于阈值电压(但接近)时,它的值比较接近串联值,而在沟道处于深度积累区时,由于很大的负栅压感应出的正电荷使表面多数载流子的浓度增加,这使它的值接近极限值C_{gc}。在深度积累的情况下,(衬底,译者注)表面处于很强的导电状态,因此基本上可以看作是一层金属,从而使栅-衬底电容完全是一个平行板电容。

电容值随偏置变化也提供了实现变容管的另一种选择。为了避免采用负电源电压,可以利用 n^+ 源区和漏区把电容制造在 n 阱内,形成一个积累模式的 MOSFET 电容。MOSFET 的各端电容如表 11.1 所示。

表 11.1 近似的 MOSFET 终端电容

	截断区	三极管(线性)区	饱和区
C_{gs}	C_{ov}	$C_{gc}/2+C_{ov}$	$2C_{gc}/3+C_{ov}$
C_{gd}	C_{ov}	$C_{gc}/2+C_{ov}+C_{ov}$	C_{ov}
C_{gb}	$C_{gc}C_{cb}/(C_{gc}+C_{cb})<C_{gb}<C_{gc}$	0	0
C_{sb}	C_{jsb}	$C_{jsb}+C_{cb}/2$	$C_{jsb}+2C_{cb}/3$
C_{db}	C_{jdb}	$C_{jdb}+C_{cb}/2$	C_{jdb}

作为最后的总结,你可以掌握一个大致的原则(如果没有给你其他信息的话),就是所有 FET 的每毫米栅宽的栅源电容为 1pF。令人惊讶的是,器件尺寸缩放的趋势并未引起该电容的变化,虽然器件特征尺寸的变化是在几个数量级,但是我们并不能保证该栅源电容保持不变的特性会一直继续下去。这个值仅仅提供零阶的粗略指导,来实现简单的估算。

渡越时间的影响(非准静态特性)

本章所讨论的集总模型显然不能用到一个任意大的频率范围上。作为一个粗略的估计原则,通常在大约 ω_T 的 1/10 或 1/5 频率以下,我们可以忽略晶体管实际的分布特性。但随着频率的提高,集总模型变得不再适用。可以认为,它最明显的缺点是忽略了渡越时间(非准静态,NQS)的影响。

为了从定性角度理解渡越时间影响的最重要的作用,考虑在栅-源之间加上一个阶跃电压,从而在沟道中感应出电荷并随之向漏极漂移,由于载流子的迁移率有限,所以在经过一段时间之后它们才到达漏极,因而跨导有一个与此相应的相位延迟。

延迟了的跨导一个伴随的效应是引起输入阻抗的变化：这个通过栅电容从沟道返回的延迟了的反馈必然会破坏栅电压和栅电流之间的完全正交关系。结果,所加的栅电压会对沟道电荷做功。任何正确的电路模型都必须要考虑这个能耗。Van der Ziel[25] 证明了至少对于长沟道器件,渡越延迟导致栅导纳出现一个实部,该实部随频率的平方而增加：

$$g_g = \frac{\omega^2 C_{gs}^2}{5 g_{d0}} \tag{26}$$

[25] *Noise in Solid State Devices and Circuits*, Wiley, New York, 1986.

为了大致确定公式(26)所代表的数值大小，假设 g_{d0} 近似等于 g_m。于是，作为一个粗略的近似，可以得到：

$$g_g \approx \frac{g_m}{5}\left(\frac{\omega}{\omega_T}\right)^2 \tag{27}$$

因此，只要保持工作在远低于 ω_T 的频率以下，这个并联电导可以忽略不计。但是为了精确计算在高频下的功率增益和热噪声，这个电导必须被包括在内。我们在以后会看到，要得到正确的噪声系数，就必须把这个噪声源考虑在内。最后，由于前面的外推单位功率增益频率推导中忽略了这些非准静态效应，因此它过高估计了 ω_{max} 的实际数值。

关于非准静态效应的最后一点，我们注意到通过将 $g_g C_{gs}$ 并联电路转换成 $R_i C_{gs}$ 串联电路，就可以用一个串联电阻代替并联电导。在这种情况下，模型参数 R_i 由下式给出：

$$R_i \approx 1/5g_{d0} \tag{28}$$

常见的一个错误是将栅源电压定义为加在电容两端的电压(在 MESFET 中，跨导器的电流的确依赖于电容电压)。但在转换为串联表示后，跨导的控制电压是一个加在 MOSFET 这个串联组合上的电压。

11.4.2 MESFET 和 HEMT 的区别

MESFET 和 HEMT 的基本小信号模型在本质上都与 MOSFET 的模型相同，虽然个别的函数依赖关系不同。尽管非准静态效应依然存在，但在 MESFET 的大多数文献中却很少提及。

对于 HEMT 的情况，栅泄漏电流通常是不可忽视的。对于这些情况，在栅电容上简单地并联一个电阻，就能使模型表征这种效应。

11.5 总结

我们已经看到，尽管可以采用更加复杂的计算机模型来进行分析，仍然可能开发出简单的小信号晶体管模型，这些简单的模型对于电路的首轮设计和作为对后续的迭代设计的指导都十分有用。即使集成电路制造商未提供合适的模型，工程师仍可采用从数据手册中提取模型参数的方法来进行工作。

第12章 放大器

12.1 引言

在微波频段的放大器设计比低频下需要更细致的考虑。当接近器件本身固有的极限时，我们只有更努力才能达到所要求的性能。另外，始终存在的寄生电容和寄生电感严重地限制了放大器所能达到的性能。实际上，由于射频频率下寄生参数是如此普遍，以至于一个重要的工程理念是利用寄生参数作为电路的元件，而不只是想方设法去消除它们。

在器件模型和仿真不够成熟的时代，传统微波放大器的设计在很大程度上忽略了器件基本的物理机理，而采用 S 参数来描述晶体管在不同频率下的宏观行为。通过这种方法可以大大简化设计，但也要付出代价。由于器件的特性与其物理机理被隔离开来，工程师很难在已有的数据基础上外推新的特性。另外，实际中使用的晶体管都是非线性的，所以 S 参数只适用于描述在给定偏置下的器件特性。

仿真和建模工具从那个时代到现在已经有了长足的发展，因此我们将从器件级观点来研究宽带和窄带放大器的设计，而不是采用传统的基于 S 参数的方法。举个例来说，我们将不会花时间来仔细研究稳定性和等增益圆[①]。对传统方法感兴趣的读者可以去参考任何一本详细阐述这方面内容的著作[②]。

我们假定读者已经熟知基本放大器结构（比如共射型，cascode 型等），因此我们将把重点放在几种拓展带宽的技术上。我们也将研究一些使放大器性能变坏的“奇怪的阻抗行为”，并讨论如何去解决这些问题。

12.2 微波偏置电路基础

我们将从实现对晶体管加偏置实用方法讲起。与低频放大器的设计不同，RF 频率下恰当的偏置方式将会受到更多的限制，因为在微波频率下很难防止偏置电路的寄生效应与信号通路相互影响。在有些情况下，可以使偏置电路起到双重作用（比如，作为匹配电路的一部分）。在所有情况下，必须对任何提出的偏置网络的物理细节很清楚。培养这个习惯对于 RF 和微波电路的初学者来说是一个挑战。很多在论文里看起来是十分有前景的电路设计，因为忽略了偏置电路的细节而达不到原来的性能指标。

因为不可能全面地去讨论各种偏置电路，我们将仅介绍几种常见的、适合于分立电路的

---

① 这里略去不讲并不是认为这种方法不好。传统的方法可以让人们在没有器件模型的条件下进行设计，这个能力是一个很大的财富。

② 比如可以参考 G. Gonzalez. *Microwave Transistor Amplifier*（微波晶体管放大器）. 2nd ed.，Prentice-Hall，New York，1996.

偏置方法。然而，这些选择的例子及其分析应该足够使读者对这个问题有一个相当完整的理解。

12.2.1　双极型晶体管的偏置

在非关键应用场合，可以使用图 12.1 所示的偏置电路。我们假设晶体管基极通过 AC 耦合到它的驱动级，那么，图上所示的元件就可以实现此电路的偏置了。这个偏置电路是保证能工作的，因为只要电源电压足够高以使得能正向导通二极管，晶体管一定能被偏置在正向有源工作区中的某一处。

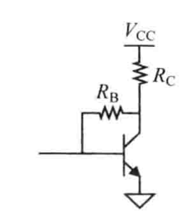


图 12.1　双极型晶体管万无一失（确保能工作）的偏置电路

最简单的证明方法便是采用反证法或称归谬法。首先，回忆一下，双极晶体管只能工作在 3 个区域：截止区、饱和区和正向有源工作区（简称有源区）。在截止区，集电结基本上没有电流。在饱和区，两个结都处于深度的正向偏置。集电极-基极电压是两个结电压的差并且比任一个的都要小。在有源区，只有基极-发射极结是正向偏置的。

现在我们有了所有需要知道的知识了。首先假设晶体管处在截止区，因此在任何支路上都没有电流，从而两个电阻上面也没有压降。基极-发射极电压等于 V_{CC}，这与晶体管处于截止区的假设是矛盾的。

作为第二个测试，假设晶体管处于饱和区。这样，集电极-发射极间电压要比基极-发射极电压小。在这种情况下，通过反馈电阻 R_B 的电流必然从基极流出。这个条件基极-发射极结处于正向偏置的要求是相矛盾，所以晶体管不可能处于饱和区。那么现在只剩下有源区一种可能了。问题就这样解决了。

这个电路另一个优点是发射极是真正接地的（而不是被交流旁路接地）。如果发射极和地之间不加入任何元件（比如电阻），寄生电感就被最小化了。这一点至关重要，因为串联电感将降低 R_F 增益，产生不想要的谐振，并且会以不希望的方式造成输入阻抗的改变。

但是与这些好特性相伴的是相对较差的偏置稳定性。集电极的电压等于基极-发射极的电压加上电阻 R_B 上的压降。而后者又取决于基极的电流，因此有对 β 的敏感性。

$$V_C = V_{BE} + \frac{I_C}{\beta} R_B \tag{1}$$

遗憾的是，β 随温度上升而增加，变化率约为 0.6%/K。因此，当温度上升时，集电极电压下降而集电极电流上升。

在图 12.2 中所分析的一个特定的偏置电路设计中，我们看到集电极电流在军事应用的温度范围内变化可达 2.6 倍。电源电压的增加会使这个变化减小，因为对于给定的 V_C 变化，当 V_{CC} 增加时其相对于在收集极负载电阻两端的电压降的百分比减少了。类似地，当设计规定的集电极电压降低时，偏置稳定性也会改善，因为落在负载上的电压提高了。

如果这些选项都不可被接受，增加一个电阻可以降低对 β 的依赖性，参见图 12.3。假设通过两个偏置电阻的电流比基极电流大。这样，加在 R_{B2} 上的电压就是 V_{BE}。如果基极电流可以被忽略，那么加在 R_{B1} 上的电压可以近似为 $V_{BE}(R_{B1}/R_{B2})$。这样，集电极电压可以近似为

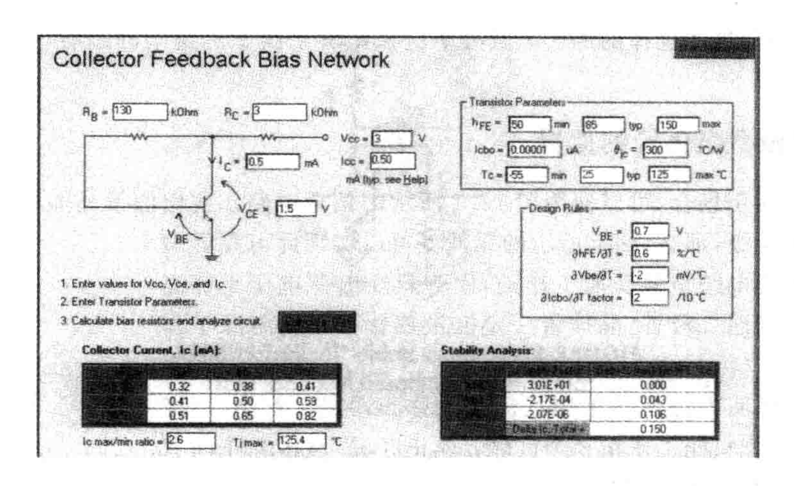


图 12.2　对于简单的万无一失的偏置电路的 AppCAD 程序模拟结果

$$V_C \approx V_{BE}\left(1+\frac{R_{B1}}{R_{B2}}\right) \tag{2}$$

输出电压，因而也就是电流，就主要取决于 V_{BE}。因为 V_{BE} 随温度的下降率约是 2mV/K，集电极电压也随温度升高而减小。在军事应用的温度范围内，V_{BE} 的改变被预测在 350mV 左右。在电源电压固定的情况下，这意味着随温度上升集电极电流将会上升，这就像我们初始提出的保证能工作的偏置电路那样。对图 12.3 这个电路采用 AppCAD 仿真，对于同样的条件，发现在这个温度变化范围内电流的波动为 2.1∶1（比原来的要小，译者注）。如果电源电压上升到 5V，随温度的波动将会降到 1.5∶1。

这种反馈偏置的方法允许发射极接到地端。偏置的稳定性来源于负反馈对 DC 增益的减小。不幸的是，这种负反馈是由宽带元件实现的，因此同样会减小对信号的增益。一个解决办法是在基极和电阻反馈网络之间插入一个阻塞电感以隔离直流偏置网络和输入端信号路径的耦合，参看图 12.4。如果 R_{B1} 和 R_{B2} 比 R_C 大得多，偏置网络对输出节点的负载效应可以忽略掉。

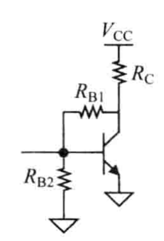


图 12.3　更为稳定的“万无一失”的双极型晶体管偏置电路（基于 V_{BE} 的反馈偏置方法）

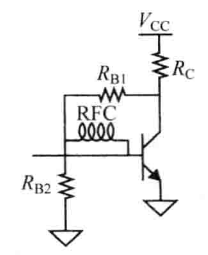


图 12.4　采用扼流圈来将 DC 与 AC 路径解耦

重要的一点是不要试图将阻塞电感设置得无限大。记住每个电感都会有寄生电容。超过自谐振频率后，电感将会表现得像电容。因此，最好采纳中庸哲学的观点，不要选择比需要还大的电感值。一个粗略的评判标准是在感兴趣频率范围的最低频率处，选择一个电抗

值比从基极看过去的阻抗大 5～10 倍的电感。

另一种在频率增高时日益吸引人的方法是用 1/4 波长传输线取代扼流电感。如果 R_{B1} 和 R_{B2} 的连接处由一个合适的电容旁路到地，那么线的另一端将表现出一个很高的阻抗。很明显，1/4 波长偏置馈线局限于一些窄带系统。如果能容忍这个限制，这种技术还是非常有用的。

还有一种方法是将反馈电阻 R_{B1} 分成两个，并将其中点旁路到地，如图 12.5 所示。当应用这种办法的时候，需要牢记的是电阻 R_{B1b} 与 R_C 相比要足够大，以不至于对输出产生过大的负载作用。同时 R_{B1a} 和 R_{B2} 必须足够大以免造成对输入端的过重负载。然而，满足这些条件有时和实现偏置的稳定性相矛盾。

和扼流电感一样，设计者必须克制使用过大电容 C 的想法。一个好的判断标准是当在放大器工作频率范围的最低频率处，选择一个电抗值比它面对的有效电阻值低 5～10 倍的电容。

来结束这个有关常见的双极晶体管偏置方法的讨论，让我们给出一种在低频下十分广泛采用的经典结构，如图 12.6 所示。因为采用了发射极简并技术，这种结构在微波频率下并不常用，除非是一些可以容忍无法避免的寄生元件所伴随的增益降低的应用场合。

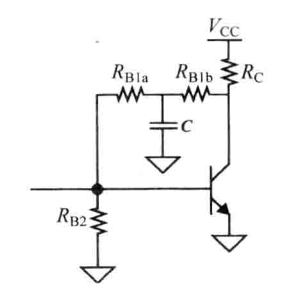


图 12.5　采用电容来解耦 DC 与 AC 路径

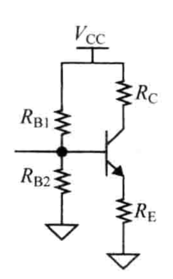


图 12.6　不依赖于 β 及 V_{BE} 的偏置电路

有很多方法可以被用来分析这种偏置结构，但是那些严格的分析方法会模糊基本的工作原理。因此，我们将提供一个近似的分析方法，这个方法对设计更为有用。与以往一样，需要用更严格的分析仿真来验证所有的设计目标是否可以满足。说明了这一点，我们将提出的方法在大部分场合下都会产生保守的结果。

偏置电流的波动来源于 β 和 V_{BE} 随温度变化而发生的变化，也因器件不同而不同。图 12.6 的电路对这两种来源都有抑制作用。为了最直接地理解电路是如何作用的，我们先考虑每个来源对偏置点的单独贡献。实际上，我们将采用小信号假设，而将两种波动简单地叠加在一起来估算总的影响。

为了估算随温度变化而造成的 V_{BE} 变化对偏置的影响，首先我们回想到，在恒定的集电极电流的情况下，V_{BE} 随温度以大约 2mV/K 的速率降低，因此在军事应用的整个温度范围内，V_{BE} 会改变 350mV。如果忽略 β 的影响，那么可以假定基极电压是固定的。这样，集电极电流的改变完全是因由 V_{BE} 改变而引起的发射极电压的变化：

$$\Delta I_{C,VBE} = \frac{\Delta V_{BE}}{R_E} \tag{3}$$

将它表示成归一化后的集电极电流的变化，可以写为

$$\frac{\Delta I_{C,VBE}}{I_{C,NOM}} = \frac{\Delta V_{BE}}{I_{C,NOM} R_E} \tag{4}$$

这样，可以通过选择足够大的 R_E 上的标称电压降来抑制集电极电流的分数变化值。例如，通过选择标称发射极电压为 3.6V，可以将集电极的电流在整个温度范围内的变化降到 10%。这个结果不应该感到奇怪，因为减小对波动的敏感度的机理是完全等同于对有用信号增益的减小。也就是说，电路实际上无法识别想要的输入信号和不想要的电压波动，因此对抑制其中之一有效的机制同样对另一种有效。

我们现在考虑如何减小对 β 的敏感度。理解 β 在不同器件（即使是名义上相同类型的器件）和不同温度之间的差别很重要。一般来说，β 随温度的变化率约为 0.6%/K。在军事应用的温度范围内，见到 β 有 2～3 倍的变化很正常。使这种变化复杂化的另一因素是工艺生产的标称分布[③]。因此，保守的设计方法是需要考虑 β 有 5 倍的变化。为了看看如何最好地做这件事，先将图 12.7 的输入电路替换为其戴维宁等效电路。这里，

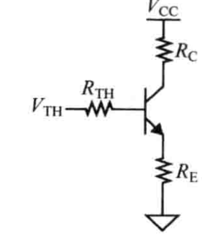


图 12.7　不依赖于 β 及 V_{BE} 偏置的等效电路

$$V_{TH} = V_{CC} \frac{R_{B2}}{R_{B1} + R_{B2}} \tag{5}$$

和

$$R_{TH} = R_{B1} \parallel R_{B2} \tag{6}$$

在分析中，我们假设 V_{BE} 是固定在其标称值上，这样集电极电流的变动来源于基极电压的变化，而这个基极电压的变化则伴随由 β 波动引起的基极电流的变化：

$$V_B = V_{TH} - \frac{I_C}{\beta} R_{TH} \tag{7}$$

很明显选择一个足够小的 R_{TH} 将可以使基极电压对 β 的敏感性降低到任意小的程度。从而也将集电极电流对 β 的敏感度降低到任意小的程度。作为一个粗略的估算法则，选择通过这些基极电阻的电流等于基极最大电流的 5～10 倍将产生令人满意的稳定度。为了提一个更加定量的指导原则，首先注意到最大集电极电流和最大 β 是一致的。这样可以写为

$$\Delta V_B = R_{TH} \left(\frac{I_{C,max}}{\beta_{max}} - \frac{I_{C,min}}{\beta_{min}} \right) \tag{8}$$

借助等式(8)，可以选择一个恰当的 R_{TH} 值来产生在功耗限制下可能达到的尽可能小的 ΔV_B。

小结一下：首先选择加在发射极简并电阻上的压降足够大来抑制基极-发射极的电压波动，然后选择基极偏置电阻足够小来抑制 β 的波动。最后，旁路简并电阻以恢复这一级的 AC 电压增益到所希望达到的值。由于寄生参数，最后一步在频率变高的时候将变得更加困难。再一次强调，一个数值很小的电感在高频下时变成一个大的阻抗。仅仅一个 1nH 的电感在 10GHz 时的值为 63Ω。

③　很少有厂商的数据手册中会详细说明这种变化，通常给出最小的 β 值而不同时给出最大值，或者简单地规定一个“典型”的值（无论这个值意味着什么）。如果你将他们的数据字面化地理解，你可能会过度设计偏置网络。

12.2.2　耗尽型 FET 的偏置

所有的结型场效应晶体管(FET)都是耗尽型器件，比如大多数 MESFET 和 HEMT。因此，在正常工作条件下，n 沟道器件的栅源电压是负的。在不追求最大效率的场合下，可以采取只使用一个简单的源简并电阻做偏置。在这种结构中，有可能将栅极接到直流地端。这种结构通常被称为自偏置(见图 12.8)，其存在可以追溯到真空电子管的年代。在实际应用中，n 沟道耗尽型 FET 是与真空管最接近的。人们通常可以很容易地将传统的真空管电路变为结型场效应晶体管电路，而仅仅需要做些很小的改动以适应工作电压的不同。

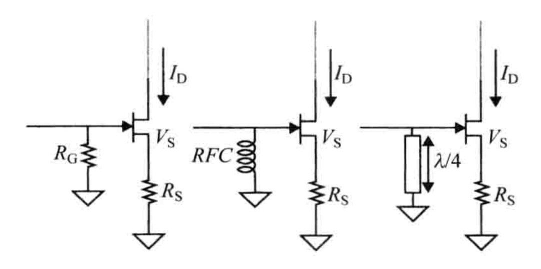


图 12.8　n 型沟道耗尽型 FET 的自偏置电路(源兼并电阻两端的旁路元件没有画出)

尽管栅极电流一般非常小，栅极端不能浮空以产生随机的电势。必须提供一个确定的直流路径来设定栅极电压，否则将会发生不少出错的状况。这里，我们假定栅极是交流耦合到前一级，这样电阻 R_G 将栅极的直流电势设为 0V。好的方面是，R_G 的值可以在一个较广的范围内选择。它的可接受的下限值要尽可能地避免信号衰减，因此它要比驱动栅极的阻抗大 5～10 倍。而它的上限值是由最坏情况下的栅极漏电流(一般发生在最高工作温度下)和最高能容忍的栅极电压的浮动而定。如图 12.8 所示，栅极电阻可以被一个扼流电感取代，或者在窄带应用中使用一个 1/4 波长的高阻抗传输线。

R_S 的值很容易从所需要的漏极电流和与之对应的 V_{GS} 计算得来的。因此，

$$R_S = \frac{|V_{GS}| \ @I_D}{I_D} \tag{9}$$

如果有从要求的漏极电流来计算 V_{GS} 的解析方程式，这会是很方便的，但并不一定要有；如果不知道器件的方程式，可以用 V-I 实验数据来得到 V_{GS}。

因为要得到的漏极电流是由器件特有的 V-I 特性控制的，因此这种偏置电路是与器件有关的。如果晶体管被替换了，同时又要求漏极电流在容差范围内，那么就要用到不同的 R_S 值。很显然，如果要求大量电路的可重复性，那么必须兼顾到偏置电路对器件特性的依赖关系。

同样重要的是注意到在这个电路中，温度导致的漏极电流波动并不是被直接控制的，设计者必须接受温度漂移带来的影响。温度变动会造成两种相反的效果：一种是随温度升高载流子迁移率的降低，这种机制将使漏极电流下降；另一种是随温度[④]上升阈值电压以

④　术语表达上的差异反映了对于增强型和耗尽型器件看法的不同。耗尽型器件在未加偏置情况下状态为导通，因此，栅电压被当作减小漏极电流的手段。最终，可以通过沟道的夹断来关断电流。然而，增强型器件在未加偏置时的状态是处于截止，要使器件导通，需要在栅极加一个有效的栅电压，即阈值电压。

－2mV/K 的速率变化(在结型场效应晶体管中这被称为关断电压)，这种效应将会增加漏极电流。对于结型场效应晶体管，漏极电流存在一个特定的点使得两种截然相反的效应在一阶近似下相互抵消，导致稳定的漏极电流。

为了方便设计与表征自偏置网络，AppCAD 提供了一种便捷的工具帮助我们确定偏置电阻的值并评估漏极电流随温度的变化。如图 12.9 所示，功能全的电路仿真器也能提供等效一样的信息，但是使用 AppCAD 的偏置设计工具更加简便。当然，它的入门指导价值是大的，因为我们很难光靠看方程式来培养设计的直觉理解，但用找个程序则可以。

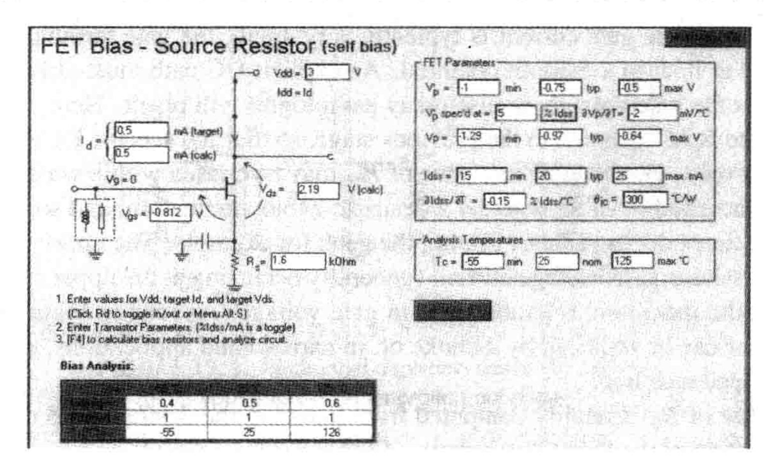


图 12.9　由 AppCAD 给出的一个样板 JFET 自偏置设计的电子表格

话虽然那么说，这里还是有一个公式：

$$I_D = K(V_{GS} - V_P)^n \tag{10}$$

式中对于传统的结型场效应管，$n=2$。常数 K 有时被称作 FET 的导流系数，因为它和描述真空电子管的 V-I 特性的类似常数相近。对于传统的平方率关系 FET，常数 K 的量纲等于电流除以电压的平方。

当我们消化数据手册时，知道参数 I_{DSS}是零 V_{GS}下的漏极电流这一点是有帮助的。另外由厂商给出的夹断电压，通常是指当漏极电流下降到 I_{DSS}的一个小的百分比时所对应的栅源电压，而不是理论上(但实验上很难测定)的漏极电流为零。这个偏置设计工具允许用户区规定这个百分比或者是给出漏极电流的绝对大小。在图中给出的例子中，这个值是 I_{DDS} 的 5%。为了将这个有些随意规定的夹断电压的定义和严格的夹断电压的定义区分开来，数据手册中给出的夹断电压是一个带撇的变量。然后，这个工具通过外推 V_{GS}-I_D 曲线到零电流来提取“真实”的 V_P。读者会注意到，这与预料中的一致，这个后者的夹断电压实际上比数据手册里给出的值更偏负一些。

从例子中用的数值(这是实际的分立器件的代表值)可以看到，漏极电流恰好是随温度升高而增加。对于这个特定器件与特定的一组数来说，最大温度下的电流比最小温度下的值大 50%。

最后，在那些对因在兼并电阻两端上的压降带来的效率或电源电压余度(headroom)的降低无法容忍的应用中，我们必须要提供负压去偏置栅极。偏置电压可以通过一个电感或电阻施加，如图 12.10 所示。在窄带应用中，另一个方案是采用 1/4 波长的传输线，正如前

面的例子。需要一个额外的电源是人们不情愿地一个十分不方便的选择。砷化镓 MESFET 曾经主宰手机的功率放大器，但它们正逐步让位于 HBT，这在很大程度上是由于后者可以在单个一电源电压下工作。

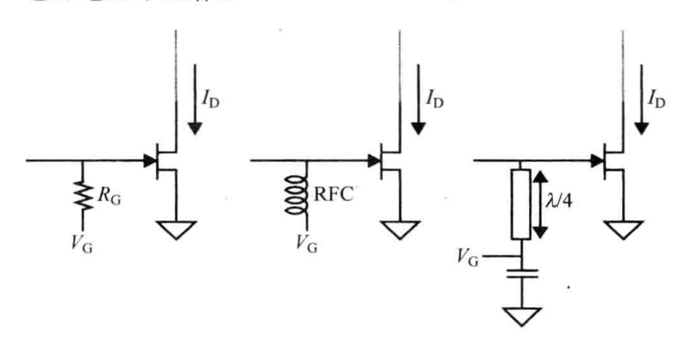


图 12.10　在额外电源($V_G<0$)具备时的对 MESFET 的一些偏置电路

还有最后一点我们需要提及：很多用于双极型晶体管的偏置拓扑结构同样可以适用于 FET 晶体管。图 12.6 电路的例子也可以用于 FET，它可以容许负的 V_{GS}。假如其中某个我们在这里给出的拓扑结构因为某种原因不适宜使用，我们鼓励读者可以去考虑替换电路。

12.2.3　有源偏置

所有前面的例子都是用无源元件去建立偏置。正如我们所看到的，只使用无源元件实现稳定的偏置不是一件容易的事情，尤其是要避免信号路径的干扰。为了使我们对偏置方法的回顾更为完整，现在来考虑增加有源器件，以提供额外的自由度。这些额外的自由度可以被用来探索比原来要精细得多的偏置控制电路。

为了验证下面要讨论的内容，假设需要放大的信号频谱没有延伸至直流。这样，仍然可以将偏置电路和信号路径分开来处理。然后，作为一种哲理性的途径，我们希望将偏置电流或电压与某个参考源比较，然后环绕放大器建立一个偏置反馈回路以强制偏置点与参考源仅有可以忽略的差别。正如可以预料到的，有很多种方法能用来去实现这个一般的想法。我们将只考虑一个方法来说明基本的原理。

一个仅依赖于增加电路复杂度的方法是采用一个运算放大器来作为有源器件。在某些情况下，这种选择可能是个很不错的主意。但是，通常对于大多数情况来说，为这个偏置的目的而使用运放是有点过分了。通常，加入一个额外的晶体管就足够了。作为一个具体的例子，考虑图 12.11 中的简单电路。

就如这里讨论的所有的其他偏置电路一样，对于这个基本的途径会有多种变形。PNP 晶体管给 NPN 晶体管提供足够的基极电流来保持 PNP 晶体管的发射极电位比它的基极高一个二极管导通电压的压降。而基极电位是由电阻分压器 R_{B1} 和 R_{B2} 实现的。这样就设定了在电阻 R 上的压降，通过电阻 R 的电流也类似地确定下来了。因为通过 PNP 器件的电流仅仅是通过 NPN 基极的

图 12.11　有源偏置的例子

电流，我们可以将其忽略，而得出结论：通过电阻 R 的电流基本上就是通过 NPN 晶体管的收集极电流：

$$I_C \approx \frac{V_{CC} - \left[V_{CC}\left(\frac{R_{B2}}{R_{B1}+R_{B2}}\right)+V_{EB,PNP}\right]}{R} \tag{11}$$

仔细研究一下式(11)，很明显这种偏置方法使得集电极电流不依赖于不可依赖的因素。然而，这并不是万无一失的。电容 C 的选择必须是足够大，并且品质要好，从而在应用的频率范围内实现射频短路。另外，我们设计了一个包含两个有源器件的反馈环路，这样不稳定的可能性增加了。如果电路的不稳定性发生了，可以去尝试降低反馈环路的增益，这可以通过在发射极增加串联电阻或者在 PNP 的收集极和 NPN 的基极之间插入一个电阻来做到。在后者情况下，通过一个电容将集电极短路到地或者使用电阻-电容串联组合(产生一个滞后补偿以改善稳定性)是有益的。不把 PNP 的集电极和 NPN 的基极直接连在一起还有一个原因，就是最好不要将 PNP 的输出电容加到 NPN 的基极节点的总电容中去。

这里所描述的各种偏置方法不可能囊括一切，但是已经足够具有代表性了。读者可以在这些基本思路的基础上设计很多其他的方法。

12.3 带宽扩展技术

12.3.1 串联和并联补偿

追溯到 20 世纪 30 年代电视机开发的初期，一个关键性的重要问题是设计具有 4MHz 视频带宽的相对平坦响应的放大器，这个带宽在当时看来是十分吃惊得宽。尽管在今天得到这样高的带宽似乎是一件十分容易的事，但考虑到那时只有真空管可以被采用，这实在是个大的挑战。在一个所用的真空管的数目就主要决定了一个电路成本的时代，工程师们设计电路时遵循的原则是可以使用任意多的无源元件，但要将用到的真空管数目严格控制至最少量。

遵循这个设计原则的一个最早期设计例子是串联成峰放大器，如图 12.12 所示。这里，我们假设电感值是唯一的设计自由度。负载电阻的值由增益指标决定，电容代表了无法减小的寄生负载，如果不采取其他措施，限制了带宽至 $1/RC$。

图 12.12 串联成峰放大器(偏置电路没有画出)

如果假设晶体管本身具有可以忽略的寄生参数，就可以把它当作一个理想的跨导器。因为我们关注所寻求的是对带宽的直观理解，可以忽略晶体管(在以上假设的范围内)而把重点仅放在输出网络的跨阻上：

$$\frac{v_{out}}{i_c} = \left[R \parallel \left(sL + \frac{1}{sC}\right)\right]\left[\frac{1/sC}{sL+1/sC}\right] = \frac{R}{s^2LC+sRC+1} \tag{12}$$

为了方便推导，将电感表达为

$$L = R^2C/m \tag{13}$$

并且使 $RC=\tau$，这是在没有采用补偿网络情况下的带宽的倒数。然后，我们将探索当 m 变

化时跨阻如何变化：

$$\frac{v_{\text{out}}}{i_{\text{c}}} = R_{\text{L}}\left[\frac{s^2\tau^2}{m} + s\tau + 1\right]^{-1} \tag{14}$$

由此式可以直截了当地证明：(a)最大带宽是未补偿情况(即 $L=0$)的两倍；(b)这个最大带宽发生在 $m=2$ 的时候。选择 $m=3$ 将实现最大的平坦化响应，相应地，带宽提升为 1.36 倍。这两种情况的带宽提升都来自于两个复极点带来的谐振成峰。

串联成峰仅需要一个电感就可以将带宽提高大约 40%。尽管这个改进已经是相当惊人了，但如果将电感移到另一个位置，还可以做得更好些。至少一直到 19 世纪 70 年代，这种并联补偿(下面会讨论到，译者注)还被广泛地应用在数不清个电视机上。它还继续时不时地出现在各种应用中，例如原始的 Macintosh 紧凑型计算机(美国苹果公司产，译者注)的视频显示电路以及用于光通信的许多宽带放大器上。

图 12.13 描述了一个最简单的并联成峰放大器。这个放大器是一个标准的共射级结构，只是集电极的电阻上增加了一个串联电感。之所以叫做并联补偿，是由于观察到电感出现在与电容并联的支路上。注意，不像串联补偿的例子，任何器件的输出电容可以被吸收到负载上，因此具有更多的使用价值。

按照我们的假设，可以将这个放大器建模为如图 12.14 所示的等效小信号电路。但在进行具体的理论推导之前，先思考一下为何按照这种方式放置电感可以比串联情况提供更大的带宽拓展。

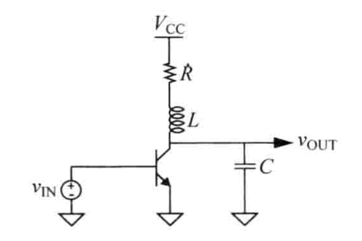


图 12.13　简化了的并联成峰放大器(偏置电路未给出)

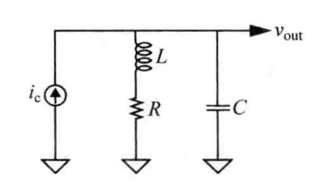


图 12.14　并联成峰放大器的模型

首先，我们知道纯电阻负载的共射级放大器的增益正比于 $g_{\text{m}}R_{\text{L}}$。我们也知道当一个电容负载加进去之后，增益最终会随频率的增高而下降因为电容的阻抗在减小。而负载电阻上串联一个电感提供了一个随频率增高而增加的成分(也就是说引入了一个零点)，它将补偿电容阻抗的减小，使得净阻抗在比原始的 RC 网络能达到的更宽的频带范围内基本不变。

通过考虑阶跃响应可以对此获得等同的时域解释。电感延时了通过含有电阻的分支的电流，使更多的电流可以用来对电容充电，从而减小了上升时间。在较快的上升时间意味着较大的带宽这个意义上，恰当地选择电感因此可以增加带宽。

将 RLC 网络的阻抗更正式地写成如下的公式：

$$Z(s) = (sL+R) \parallel \frac{1}{sC} = \frac{R[s(L/R)+1]}{s^2LC + sRC + 1} \tag{15}$$

我们看到就像在串联成峰的情形一样，有两个极点(可能为复数)。此外，正如早先所分析的，的确还有一个零点。

由于该放大器的增益是 g_{m} 和 $Z(s)$ 幅度的乘积，现在让我们来计算后者和频率之间的

函数关系：

$$|Z(j\omega)| = R\sqrt{\frac{[\omega(L/R)]^2+1}{(1-\omega^2 LC)^2+(\omega RC)^2}} \tag{16}$$

和串联成峰一样，现在引入一个因子 m，把它定义为时间常数 RC 和 L/R 的比例，以便简化后面的推导：

$$m = \frac{RC}{L/R} \tag{17}$$

于是，转移函数变为

$$Z(s) = (sL+R) \parallel \frac{1}{sC} = R\frac{\tau s+1}{s^2\tau^2 m + s\tau m + 1} \tag{18}$$

其中 $\tau = L/R$。

我们在前面没有进行串联补偿最优条件的细节推导，因为并联成峰电路性能更好（我们将马上看到）。因为了解这个推导过程是有启发意义的，让我们现在更仔细地研究这个并联成峰放大器。

用 DC 值（$=R$）来进行归一化的与频率有关的阻抗幅度为

$$\frac{|Z(j\omega)|}{R} = \sqrt{\frac{(\omega\tau)^2+1}{(1-\omega^2\tau^2 m)^2+(\omega\tau m)^2}} \tag{19}$$

所以

$$\frac{\omega}{\omega_1} = \sqrt{(-m^2/2+m+1)+\sqrt{(-m^2/2+m+1)^2+m^2}} \tag{20}$$

式中，ω_1 是未经补偿的 -3dB 频率（$=1/RC$）。

目标是决定一个 m 的值使得我们能够得到某些所希望的特性。正如在串联成峰的情况，我们有不止一种的选择来决定，使带宽最大显然是一种自然的选择。对表达式(20)求微分，将微分取为零，经过相当程度的工作，我们发现这个最大值发生在如下的 m 值处：

$$m = \sqrt{2} \approx 1.41 \tag{21}$$

上式使带宽扩展为未加补偿时的 1.85 倍。任何曾努力去满足难以实现的带宽要求的人都会体会到，通过只增加一个电感使带宽几乎翻倍而不增加功耗是一件很成功的事情。

然而，遗憾的是，m 的这一选择导致了频率响应产生了几乎 20%的过冲峰值，这样大的峰值通常是被认为是不可接受地高，见图 12.15。为了控制这个峰值不至于太高，可以通过增加 m 来寻求一个不同于绝对最大的带宽。一个通常的选择就是设频率等于未补偿的带宽时阻抗的值等于 R。求解这个条件得到 m 值等于 2，此时响应的带宽为

$$\omega = \omega_1\sqrt{1+\sqrt{5}} \approx 1.8\omega_1 \tag{22}$$

因此，在这种情形下带宽仍非常接近最大值。进一步的计算表明频率响应峰值被显著地减小到大约 3%。

得到这个结果的这种人为选择常被采用，因为它显著地提高了带宽而不产生过多的频率响应过冲峰值。然而，在许多其他情形中人们希望频率响应完全不出现峰值，因此我们也许可以寻找一个 m 值使带宽最大且满足没有任何峰值的限制。

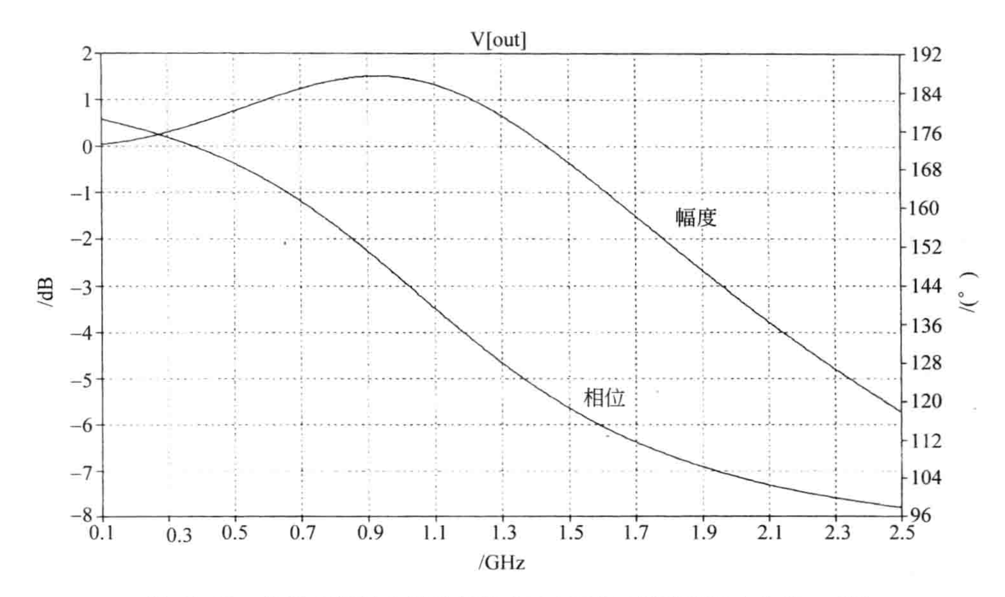


图 12.15　并联成峰放大器的频率响应(对应于使带宽最大化的 m 值)

得到这种最大平坦响应的条件可以通过以下的一般方法来得到：通过一个频率响应幅度(或常常更为方便的，是幅度的平方)的表达式，然后使在 DC 时导数为零的各阶导数数目最多。

手工来运用这个方法通常是十分费工的，但对这个具体例子，直接计算得到 m 这个奇妙的值为

$$m = 1+\sqrt{2} \approx 2.41 \tag{23}$$

上式使带宽扩展为没有峰值时的 1.72 倍。因此，至少对于并联成峰放大器而言，可以同时得到最大程度的平坦响应度和显著扩展的频带宽度。

在其他情形中，也许是对放大器的时域响应而不是它的频率响应有专门的要求。具有实际兴趣的一个例子是超宽带(UWB)系统，那里许多系统要求好的脉冲保真度。另外一个例子是示波器的垂直偏转放大器，它的时域响应(是用阶跃和脉冲响应来表征的)必须类似地有好的“行为”。也就是说，我们不仅要求在一个实际尽可能大的带宽上均匀地放大信号的各个频谱成分，而且要使其傅里叶分量之间的相位关系也必须保持不变。如果各个频谱成分经历的延迟(用绝对时间而不是角度来衡量)不相同，那么严重的波形失真就可能会发生。这样的“相位失真”是有害的，因为它能在数字系统中引起误码或者对像示波器这样的模拟信号仪器的保真度产生明显的负面影响。

为了使这类失真定量化，首先考虑一个纯时间延迟的相位特性。如果所有的频率都延迟一个相同数量的时间，那么当频率增加时这一固定数量的时间延迟必然代表线性增加的相移量。如果与这个理想线性相移的偏差最小，那么相位失真也将最小。

因此很明显，我们希望考察延时作为频率的函数。如果这个延时对于所有频率都是相同的，那么我们将不会有任何相位失真(除了波形的变化，这是由任何具有有限频带的放大器提供的通常的滤波作用而引起的)。形式上，这个延时定义为

$$T_{\mathrm{D}}(\omega) \stackrel{\mathrm{def}}{=} -\frac{\mathrm{d}\phi}{\mathrm{d}\omega} \tag{24}$$

式中，ϕ 是频率为 ω 时放大器的相移。

遗憾的是，一个有限阶次的网络不可能在一个无限大的带宽上提供常数的时间延时，这是因为这会最终要求有无限大的相移，而极点和零点只能贡献有限大小的相移。因此，我们实际上所能做的就是在某个有限的带宽上提供近似的常数延时。

与频率响应的情形类似，可以看到如果我们使在 DC 时为零的 $T_D(\omega)$ 导数的数目最多，那么就可以得到最大平坦的时间延迟。这里需要再次说明，这个方法是具有一般性的。

由于极点和零点的相移表达式中涉及到反正切函数，因此计算相关的导数比起计算复数模的情形来说，一般要更麻烦一些。即便对于我们的仅仅是二阶的并联成峰放大器，计算的工作量也是相当大的。然而，最终我们可以推导出有关 m 的三次方程（采用计算工具在这里会有极大的帮助）：

$$m^3 - 3m^2 - 1 = 0 \tag{25}$$

其相关的根为

$$m = 1 + [(3+\sqrt{5})/2]^{1/3} + [(3-\sqrt{5})/2]^{1/3} \approx 3.104 \tag{26}$$

这相应于带宽改善的倍数比 1.6 略低一点。

图 12.16 和图 12.17 所示的两个图使得我们能估算最大平坦化延迟，这是通过比较两个并联成峰放大器的脉冲响应来估算得到，一个是设计成有最大带宽，另一个则是有最大线性相移。看上去十分明显，两个脉冲响应具有类似的上升和下降时间，这意味着具有相似的带宽。然而，线性相移这个设计中的脉冲响应看起来要好得多。

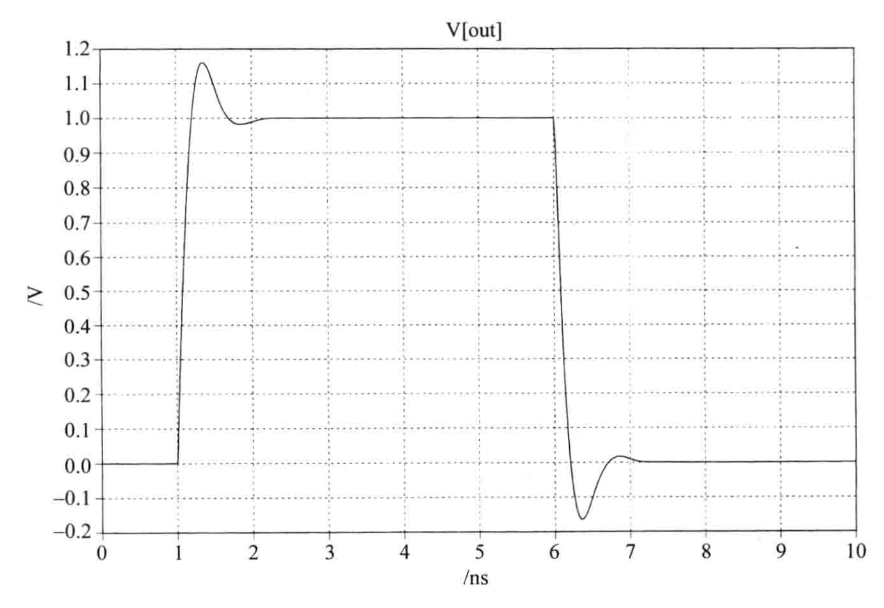


图 12.16　并联成峰放大器的脉冲响应（m 的选择使得带宽最大）

但由于达到最平坦频率响应的条件与达到最大平坦时间的时延的条件并不一致，就像在串联成峰放大器的例子中一样，我们必须进行折中选择。因此看到根据不同的要求存在着一个有用电感值的范围，见表 12.1。较大的 L（即较小的 m）导致较大的带宽扩展，但脉冲响应的保真度较差；而较小的 L 对带宽的改善较少，但却具有较好的脉冲响应。

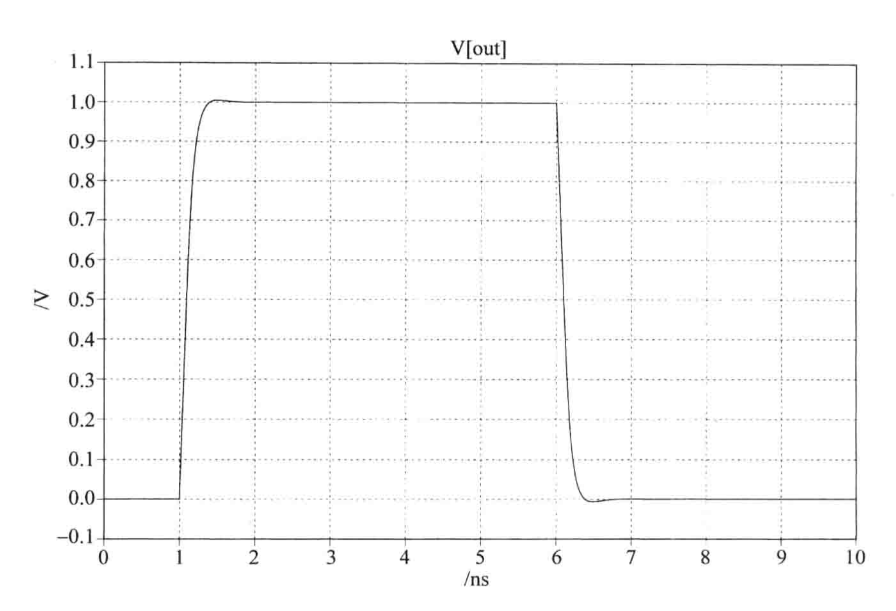


图 12.17　并联成峰放大器的脉冲响应(m 的选择使得延迟最为平坦)

表 12.1　并联成峰的小结

条　　件	$m=R^2C/L$(近似)	带宽增加因子	归一化的峰值频率响应
最大带宽	1.41	1.85	1.19
$\|Z\|=R@\omega=1/RC$	2	1.8	1.03
最佳幅度平坦度	2.41	1.72	1
最佳延迟平坦度	3.1	1.6	1
没有并联成峰	无穷大	1	1

并联成峰补偿的例子

尽管并联成峰补偿起源可以追溯至 20 世纪 30 年代以来的视频放大器，但出于与其最初被发现有价值的相同原因，它在现代依然是一种有用与被使用的技术。该方法使得我们可以从一个给定的工艺技术中最大程度地挖掘额外的性能。这种技术并不要求高 Q 值的成峰电感，因而容易被实现。注意到这些很有用。为了进一步说明这一点，考虑设计一个 1.5GHz 的共射宽带放大器的问题。让我们假设对这个应用而言，相位的线性性是很重要的，所以选择 $m=3.1$ 来得到最佳的群延时一致性。

假设(图 12.18 中的，译者注)集电极上的总电容负载为 1.5pF(包括晶体管本身的以及由互联线及下一级造成的负载效应)，并假设负载电阻不能比 100Ω 更小，否则为使增益保持为常数会使功耗增加一个不可接受的量。如果整级放大器的带宽完全是由输出节点决定的，那么这个放大器的带宽就只比 1GHz 稍大一点，还达不到 1.5GHz 的目标。

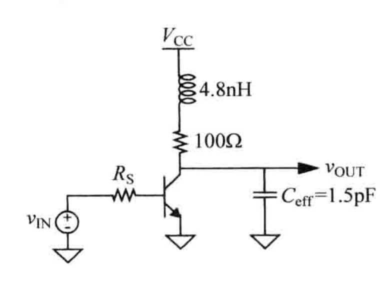


图 12.18　并联成峰放大器的例子

如果假设采用最小的可接受的电阻(即 100Ω)，那么可以很容易地计算出所要求的并联补偿电感如下：

$$L \approx R^2C/3.1 = 4.8\text{nH} \tag{27}$$

这个电感值在许多技术中都容易被实现，无论是用分立元件还是集成的，如图 12.18 所示。这样估计出来的带宽将增加到大约 1.7GHz，很容易地就超过了所要求的 1.5GHz。再一次指出，这一带宽的改善并没有增加这一级的功耗。最后，我们注意到集电极网络的 Q 值在 1.7GHz（带宽边缘）时约为 1，所以采用有一般 Q 值的电感（如 IC 中的螺旋线电感）就能满足要求。任何与电感串联的电阻又形成了电阻负载的一部分，更加使得电感的损耗不是那么地重要了。

从图 12.19 给出的仿真结果容易看出，带宽已经被提升到了比 1.7GHz 稍稍低一点。线性的频率轴让我们容易看到相位和频率保持线性增长关系，直到频率接近放大器的带宽，意味着在放大器的通带内有一个相对恒定的时延。

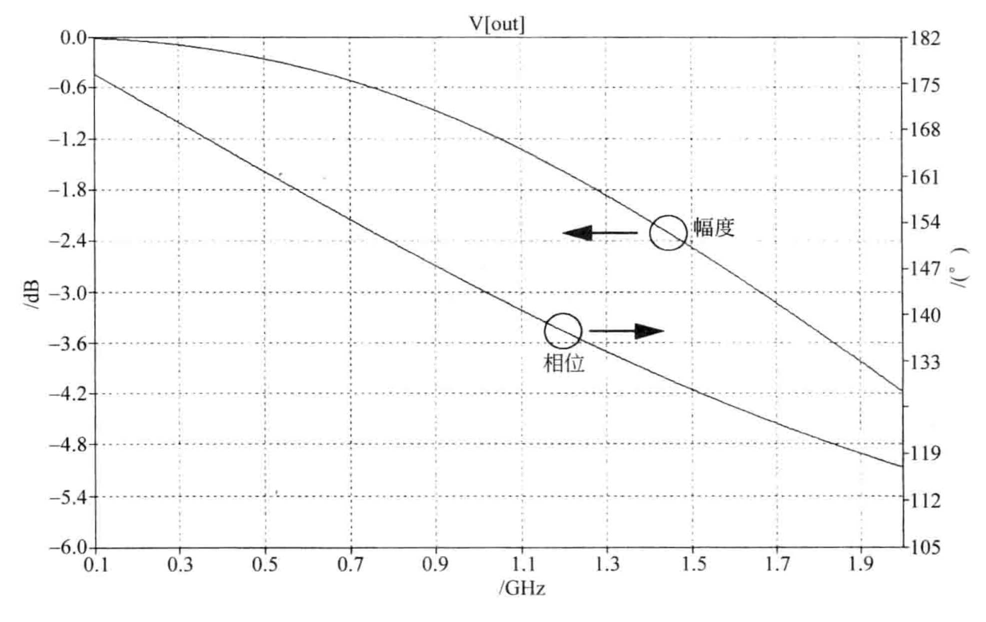


图 12.19 并联成峰放大器的 LTSpice 程序模拟结果（$m=3.1$ 以取得最佳的延迟）

12.3.2 更多关于用零点来拓展带宽

并联成峰补偿放大器演示了零点的用处。将零点想成一个反极点可以帮助我们理解它为何能拓展带宽。然而，有个问题是零点是无法精确地和它所要抵消的极点的频率完全相等的，所以我们借助下面这个传输函数来评估这两个频率不完全相等所造成的影响：

$$H(s) = \frac{\alpha\tau s + 1}{\tau s + 1} \tag{28}$$

为了简单起见，这个式子里没有加入增益或衰减因子。因此，常数 α 的理想值是 1，这样 $H(s)$ 在所有频率处都是理想的单位值 1。

让我们考虑这个组合（通常被称作极点-零点对）的阶跃响应。初值和终值定理告诉我们初值是 α 而终值是 1。因为系统状态随极点所对应的时间常数指数型地演变⑤，所以我们

⑤ 由于这个说法理解起来不那么明显的原因，关于这点可能有相当多的疑问。零点的出现仅仅是改变初始误差，但是这个误差经过一个极点的时间常数后，总是稳定于终值决定的值。

可以很快地画出一系列可能的阶跃响应：一个是 $\alpha<1$，一个是 $\alpha>1$，见图 12.20。

我们看到响应马上跳变到 α，但是然后在极点的时间常数内慢慢稳定到终值。如果 α 正好等于单位值，那么响应达到终值不需要时间。当然在所有实际电路中，额外的极点将导致上升时间不为零，但是从这个例子来看这个一般性思想应该是很清楚的。同时我们也看到带宽的拓展并非是完全没有代价的，因为详细的时域响应并不是一个真正的阶跃。

这种抵消可以用图 12.21 所示的简并共射放大器来实现。这里，C_E 并不是取得大到使在全部工作频率范围内都形成短路。实际上，它的值只需大到当 C 开始短路 R 时，它开始短路 R_E 即可。因此容易理解理想的补偿应该在 $R_EC_E \approx RC$ 的时候实现。为获得最佳的响应，还有必要作适当的调整。

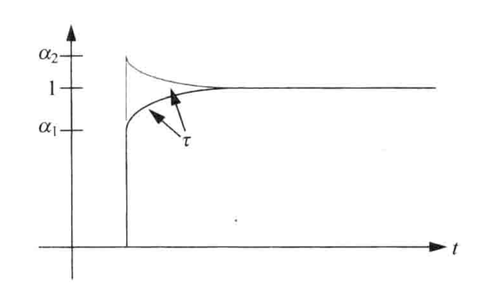


图 12.20　极点-零点对的可能的阶跃响应

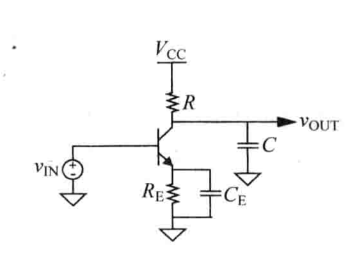


图 12.21　利用零点成峰的共发放大器

12.3.3　二端口网络的带宽拓展

并联峰值是一种在放大器和负载间连接一个一端口网络的改善带宽的方式。串联峰值则是在放大器和负载间插入一个二端口网络。尽管并联峰值技术比串联峰值能提供更大的带宽改善，但不能得出结论说二端口改善技术就比较差（尤其是因为一端口网络实际是二端口网络中的一个特例）。

我们可以通过将负载电容和器件的输出电容分离开来增强并联峰值。如果使用一个串联电感来实现这种分离，那么最终的结果是并联峰值和串联峰值的组合，见图 12.22。我们不会花太多时间分析这个组合，因为它只是一个中间步骤。下一步演变性的步骤（也是一个中间步骤）是在器件和网络的其他部分之间加一个电感，如图 12.23 所示。

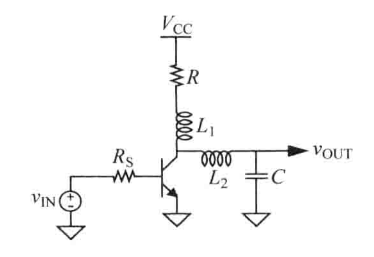


图 12.22　采用并联与串联成峰的放大器

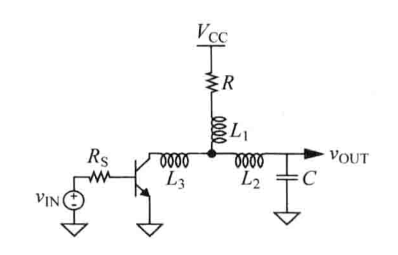


图 12.23　并联与(双)串联成峰放大器

如果我们忽略晶体管的输出电容(因为 L_3 是和电流源串联在一起的)，这种组合在功能上和上一步相同。这样让我们再定性地考虑这个问题。正如对一个普通的并联成峰放大器的阶跃响应一样，流入负载电阻的电流由于 L_1 的影响而延迟。这个行为单独就可以加速对负载电容的充电。除了这个机制外，晶体管在最初一段时间内只驱动它自身的输出电容(没有画出)，这是因为 L_3 延迟了流往网络其他部分的电流。因此集电极的上升时间改善了，这我们也可以解释成对带宽的改善。在集电极电压上升到相当高之后的一段时间后，负载电容上的电压也开始随着电流流入 L_2 而上升。因此，这个网络在时间上顺序而不是并联地对电容充电。折中是以延时为代价来改善带宽。我们将会看到带宽-延时折中是一个反复发生的情况。

为了节省空间，三个电感的组合可以通过一对磁耦合的电感(也就是变压器)来实现，因为这种连接的等效电路正好是我们想要的(考虑一下变压器的 T 模型)。也就是说，电感 L_2 可以白白得到(而且其值根据通过两个线圈的相对电流方向，可以取正值或负值，这个网络需要一个负的值)。这个电感可以由一对放置足够近的线圈来产生所想要的耦合。以 Tektronix 的 500 系列示波器为例，这个电感是由一个塑料杆上缠绕两个单层线圈构成的。将两个线圈之间的距离适当安排，即可获得想要的耦合因子。

在一个 $L_1=L_3=L$ 的专门情形下，我们可以推导出负载网络的跨阻为

$$\frac{v_{\text{out}}}{i_{\text{c}}}=R\frac{s(L/R)+1}{s^2(L+L_2)C+sRC+1} \tag{29}$$

注意 T 形线圈网络的传输函数和并联成峰的实际上是一样的，都包含两个复极点和一个零点。正如前面提到的，我们如果忽略晶体管的输出电容，这个网络相比于并联成峰补偿电路是没有更多的优点的。

在电感两端跨接一个小的电容创造一个并联谐振可以提供进一步性能改善。谐振所引起的增加了的循环电流可以帮助带宽的再提升。推导这一情形下的跨阻过程直接但非常繁琐，所以我们将详细推导过程推迟放到附录(12.8 节)中去。正如在那里所示，这个繁长的推导得到的结果是

$$\frac{v_{\text{out}}}{i_{\text{c}}}=R\frac{2s^2LC+s(L/R)+1}{s^4L(L+2L_2)CC_{\text{B}}+s^3 2RCLC_{\text{B}}+s^2[C_{\text{B}}(L+L_2)+2LC]+sRC+1} \tag{30}$$

经历了比并联成峰电路漫长的推导过程更难熬的焦躁之后，我们可以证明这对耦合的电感每一个有以下的值：

$$L_1=L_3=L=\frac{R^2C}{2(1+|k|)} \tag{31}$$

其中，L 可以理解为当另外一边开路时的初级或次级线圈的电感(完整的推导见附录)。因此，比如说，这就是在 IC 实现中对每个螺旋电感设计与画版图时的值，或者在低频工作时用分立元件实现时每一个单层螺旋管的值。

跨接电容的值为

$$C_{\text{B}}=\frac{C}{4}\frac{1-|k|}{1+|k|} \tag{32}$$

在附录里也证明了一个耦合因子为 1/3 会产生巴特沃斯型(最大幅度平坦)响应，而 $k=1/2$ 会导致最大平坦群延时响应。这些耦合因子都不算太大，因此实现起来比较容易。比如，两个挨得很近的键合线电感的耦合系数就在这个范围内。

应用这些条件就得到了图 12.24 所示的电路。最终的网络被称作桥接 T 形线圈，它在示波器电路里已经被使用了将近 40 年[⑥]。在器件的输出电容相对于负载电容小得多的条件下，这个电路几乎可以将带宽拓宽 3 倍(实际可达到的最大改进是 $2\sqrt{2}$，或者说 2.83 倍，这个值是在巴特沃斯条件下得到的)。

可以证明，如果两个电感的结合点驱动具有较高电容节点的话，那么带宽可以最大化。在图 12.24 中，我们假设负载电容比晶体管的输出电容大(记得，我们经常假设晶体管的输出电容为零)。如果输出电容比负载电容大的话，集电极和负载电容的连接法可能要反过来。

作为最后的一点改进，通过与输出电容串联一个电感可以提供额外的补偿，这样做等效提供更多的串联成峰。一个几乎相同的结果可以通过不在电感的中点，而是在线圈其他点引出抽头(在这个例子里，抽头点要更靠近负载电阻)来得到。

修改我们早前的并联成峰的设计例子得到图 12.25 的电路。这里，为了比较公平起见，继续假设希望达到的是最大群延时平坦化。正如所看到的，总的电感被翻了 1 倍。然而，这两个电感可以有一定的错移叠放在一起，增加的面积并不大(大约 50%)。另外，125fF 的跨接电容可以由两个电感之间存在的寄生电容提供。这个电路提供的理论带宽改善因子在 2.7 左右[⑦]。因此，可以期待有 2.9GHz 的带宽，比并联成峰值放大器的 1.7GHz 带宽要好得多。这种改善是在不需要增加功率，也不需要更先进的器件技术情况下取得的，强调这一点很重要。

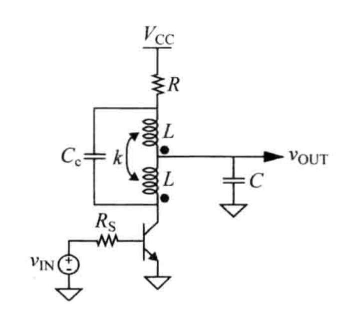


图 12.24　采用有跨接电容的 T 形线圈来增加带宽的放大器

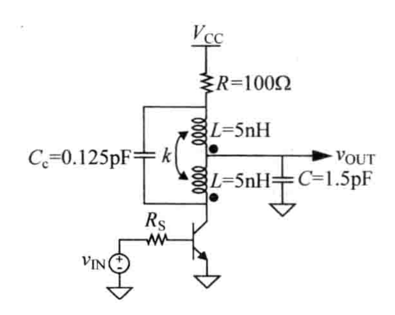


图 12.25　电容跨接 T 线圈带宽拓宽电路的例子

图 12.26 的仿真结果表明，这个电容跨接 T 线圈网络的确提供了期望的带宽提升效果。而且，线性的频率轴使得我们可以从相位图看出直到放大器的带宽都有线性的相位特性(意味着时延是常数)。

尔后我们会赞赏图 12.25 所示的结构，因为它本身是演变成完全的分布式放大器的一个中间步骤。在这个电路中，寄生电容被吸收到结构中去，用延时而不是带宽来换取增益。

⑥　然而在 Tektronix 500 和 7000 系列示波器生产的整个时期内，这些设计公式是商业秘密。这些电路网络被给出时，并没有设计公式，而只有最粗略的定性描述(Bob Orwiler's *Oscillascope Vertical Amplifiers*(示波器的垂直放大器)，Tektronix Circuit Concepts Series，Tektronix，Beaverton，OR，1969)。

⑦　再次说明一下，这个值是在假设晶体管的输出电容和负载电容相比可以小到忽略的情况下得到的。如果这个不等式不能很好地满足，则需要额外的串联补偿来达到这个数量级的带宽拓展。

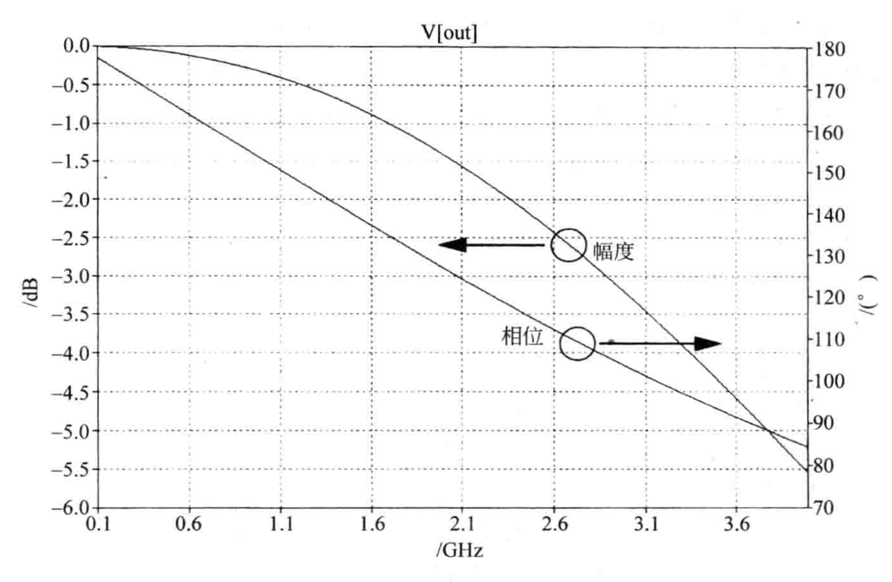


图 12.26　电容跨接的 T 线圈放大器的 LTSpice 模拟结果(线性相位情况)

(考虑一个传输线,它包含电感和电容,但是这些元件对带宽没有限制,因为电容在时间上是依次被充电的。)同时,这个电路也可以简单地被看作是用一个更为复杂的手段来将流往负载电阻的电流转流到负载电容上去。

12.4　并串放大器

不同于我们至今已经介绍过的开环结构,设计宽带放大器的另一种方法是采用负反馈。一种特别有用的采用负反馈的宽带电路就是并联-串联放大器。其名字来自于它采用了并联和串联反馈的组合,它的用途在于其输入和输出阻抗在一个很宽的频带范围内相对恒定且相同(这可以使级联大大简化)以及从这种电路设计起来比较容易。此外,双反馈环路也提供了通常负反馈所能提供的益处,也就是降低了对于器件参数的依赖性、改善了失真、扩展了带宽,而且拓扑结构看起来也顺眼。

去掉偏置细节的并联-串联放大器如图 12.27 所示。图中,R_S 是输入信号源的电阻,R_L 是负载电阻。因此,该放大器的核心部分只包括 R_F,R_E 及晶体管。为了理解这个放大器是如何工作的,我们最初假设 R_E 足够大(相对于晶体管 g_m 的倒数),所以它使总的跨导减小到接近 $1/R_E$。由于 R_E 与输入和输出回路都相串联,所以因 R_E 引起负反馈是这个放大器名字中串联这部分的来源。为了进行进一步分析,同样也假设 R_F 足够大以至于它对于输出节点的负载效应可以忽略。在这些假设条件下,这个放大器从基极至集电极的电压增益近似为 $-R_L/R_E$。

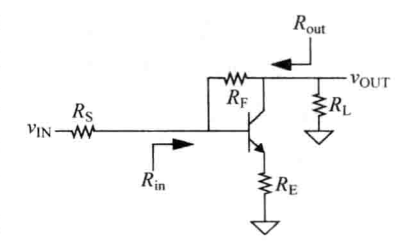


图 12.27　并联-串联放大器(偏置电路未给出)

尽管我们已经假设 R_F 对增益只有很小的影响,但它对输入和输出电阻具有控制性的

作用。特别是通过它所提供的(并联)反馈使这两个电阻同时减小。此外,由于输入和输出电阻的减小也减小了开路时间常数的和[8],从而有助于带宽的进一步提高。

为了计算输入电阻 R_{in},我们利用了从基极至集电极的增益接近 $-R_L/R_E$ 这个事实。如果我们可以忽略基极电流(这看上去十分合理),那么输入电阻完全来自流经 R_F 的电流。当在基极加上一个测试电压源时,就可以用通常的方法来计算等效电阻。正如在经典的米勒效应中那样,把一个阻抗连在存在反相位增益的两个节点之间将使阻抗减小。写成公式时,R_{in} 可以表示为

$$R_{in} = \frac{R_F}{1 - A_V} \approx \frac{R_F}{1 + R_L/R_E} \tag{33}$$

式中,A_V 是从基极到集电极的电压增益。

现在,为了计算输出电阻,在集电极节点上加一个测试电压源并再次取 v_{test} 与 i_{test} 的比:

$$R_{out} = \frac{R_F + R_S}{1 + R_S/R_1} \approx \frac{R_F}{1 + R_S/R_E} \tag{34}$$

如果电源电阻和负载电阻相等(这在分立元件的实现中是一个特别普遍的情形),那么公式(33)和式(34)的分母就大致相等。由于分子也大致相等,所以可以得到 R_{in} 和 R_{out} 它们也大致相等。如果 $R_S = R_L = R$,则可以写出

$$R_{out} \approx R_{in} \approx \frac{R_F}{1 + R/R_E} \approx \frac{R_F}{1 - A_V} \tag{35}$$

这个放大器能够很容易地同时在输入端口和输出端口处实现阻抗匹配是它被普遍采用的原因之一。一旦选定阻抗大小和增益,就能很容易确定所要求的反馈电阻值。根据已知的负载电阻与所要求的增益,可以很快求出所需要的 R_E 值。为了完成这个设计,还必须选择合适的器件宽度及偏置点。一般来说,这些选择需要保证有足够的 g_m 以使得到上面一组公式时所做的假设成立。

12.4.1 并联-串联放大器的详细设计

前面的介绍概述了并联-串联放大器的一阶特性以帮助建立设计的直观感觉。然而为了进行更加详细的设计,我们现在考虑一些前面一节中被忽略的二阶效应。

低频增益和输入-输出电阻

首先计算从基极至集电极的增益,因为它使我们可以很容易地发现输入和输出电阻。一旦已知基极至集电极的增益和输出电阻,总的增益就可以非常容易地从分压器的关系中求出。

首先回想一下一个带有发射极简并电阻的共发射极放大器的有效跨导为

$$g_{m,eff} = \frac{g_m}{1 + g_m R_E} \tag{36}$$

注意,由公式可知:只要 $g_m R$ 比 1 大许多,等效的跨导就可以近似为 $1/R_E$.

在基极和地之间加一个测试电压时,集电极电流将同时流过负载和反馈电阻。测试电压的一部分也会直接前馈至输出端。叠加原理使我们可以分别处理构成输出电压的每一部分电压:

---

[8] 参见 T. H. Lee. CMOS 射频集成电路设计. 第二版,第八章,剑桥大学出版社,2004。

$$v_{out} = -g_{m,eff} v_{test} \frac{R_F R_L}{R_F + R_L} + v_{test} \frac{R_L}{R_F + R_L} \tag{37}$$

对增益求解得到

$$A_V = \frac{v_{out}}{v_{test}} = -\frac{R_L}{R_E} \cdot \left[\frac{1}{1 + 1/g_m R_E}\right] \cdot \left[\frac{1}{1 + R_L/R_F}\right] \cdot \left[1 - \frac{1}{g_{m,eff} R_F}\right] \tag{38}$$

公式(38)虽然不是以最简洁的形式出现，但它表明从前面的一阶原理推导出的增益要乘以 3 个因子(都是在方括号内)，每个因子在理想情况下都变为 1。

第一个非理想因子反映了有限大的 g_m 对于有效跨导的影响。虽然在 $g_m R_E$ 很大的极限情况下 $g_{m,eff}$ 接近 $1/R_E$，这第一个因子从定量上表现了 $g_m R_E$ 值为非无穷大时的影响。第二项来自于 R_F 对输出节点的负载效应。只要 R_F 比负载电阻 R_L 大很多，那么因为这个因子而引起的增益降低是很少的。

最后一项降低增益的因子是由于输入信号前馈至输出端。这个前馈之所以降低增益，是因为原先的增益路径是反相位的，而前馈路径却不是反相位。因此，前馈路径部分抵消了所希望的输出。前馈部分的跨导为 $1/R_F$，因此只要这个寄生跨导与所希望的跨导 $g_{m,eff}$ 相比很小，那么这个增益损失就可以忽略。

在逐项考察了这个完整的增益公式后，我们现在提出一个简洁得多的(但仍然是准确的)表达式，它对于后面的计算式很有用：

$$A_V = -\frac{R_L}{R_{eff}} \cdot \left[\frac{R_F - R_{eff}}{R_F + R_L}\right] \tag{39}$$

式中，R_{eff} 就是有效跨导的倒数。其结论非常简单：为了得到所希望的增益，必须选择 R_E 或 R_{eff} 的值，使它比根据一阶公式所预期的值小一些。

现在我们有了一个(甚至是两个)低频增益的完整表达式，因此就可以得到在基极与地之间电阻的更为精确的值：

$$R_{in} = \frac{R_F}{1 - A_V} \tag{40}$$

利用公式(39)，上式变为

$$R_{in} = \frac{R_F}{1 + \frac{R_L}{R_{eff}}\left(\frac{R_F - R_{eff}}{R_F + R_L}\right)} = \frac{R_{eff}(R_F + R_L)}{R_{eff} + R_L} \tag{41}$$

一般来说，我们的设计都是想得到一个特定的增益值。假设能够成功地达到这个目的，那么为得到所希望的输入电阻所需的反馈电阻的值就可以很容易地用公式(40)求出。

输出电阻(即是，为 R_L 所看到的)也很容易求出。同样，我们把一个测试电压加到集电极节点上并且计算测试电压与测试电流的比。这一步完成后得到：

$$R_{out} = \frac{v_{test}}{i_{test}} = \frac{R_F + R_S}{1 + g_{m,eff} R_S} = \frac{R_F + R_S}{1 + R_S/R_{eff}} = \frac{R_{eff}(R_F + R_S)}{R_{eff} + R_S} \tag{42}$$

比较这些输入和输出电阻的表达式可以看到，如果 R_S 和 R_L 相等(如通常情形中那样)，那么 R_{in} 和 R_{out} 也将精确相等。这一令人愉快的偶然一致性正是这种放大器拓扑结构被极为广泛采用的一个原因[9]。

---

⑨ 需要指出，存在着一个小小的不同，即有限的 β 值会引起输出和输入电阻有一点不相等，即使对典型的 β 值而言，这个误差是很小的。输入电阻会有大约 $1-1/(2\beta)$ 的倍数而变得小些，而输出电阻会有大约为 $1+1/(2\beta)$ 的因子而变得大些。增益也会略有下降，减小的倍数大约为 $1-1/(2\beta)$。

应当强调的是，当进行设计时（这与分析正好相反），所要得到的增益是已知的。因此，如果输出和输入电阻是要设计成相等，选定反馈电阻的值由公式(40)是十分容易的。然后，就可以选择 R_E 的值来得到正确的增益，这样就完成了设计。

带宽和输入-输出阻抗

在介绍了对各种低频时的参量（增益与输入-输出电阻）的准确表达式后，我们现在来推导这个放大器的带宽以及输入输出阻抗的表达式。

在埋头钻进大量公式之前，让我们看一下是否能预见这些量的定性特性。由于这个放大器是一个低阶系统，所以可以期待在增益和带宽之间的折中或多或少是线性的。另一方面，正是由于它是一个低阶系统，所以采用开路时间常数来估计带宽应该是相当准确的。

我们也可以预见输入阻抗中包含一个电容成分，这部分既是由于 C_{gs} 的存在，也是由于密勒效应对 C_{gd} 的增强效应（译者注：作者在这一节讨论的似乎是针对 MOSFET 进行的，例如 C_{gd} 隐含着指栅漏间的电容，而不是基极与集电极间的电容。所幸的是，这些讨论对于 BJT 与 FET 都是适用的）。另一方面，输出阻抗的行为则有所不同，这是因为并联反馈会使输出电阻降低，而这个机理随频率的提高而变得不是很有效。结果是，输出阻抗实际上可能随频率而上升，导致了输出阻抗中包含有一个电感成分。

在做了这些预测之后，让我们继续计算开路时间常数的和。为了简化推导过程，假设器件只有电容 C_{gs} 和 C_{gd}，而且忽略串联的基极电阻。最后假设电源内阻和负载电阻的值都等于 R。

很显然，面向 C_{gd} 的有效电阻是 R_F 与下面给出的电阻的并联：

$$r_{\text{left}} + r_{\text{right}} + g_{m,\text{eff}} r_{\text{left}} r_{\text{right}} \tag{43}$$

所以有效电阻为

$$R_F \parallel (R_S + R_L + g_{m,\text{eff}} R_S R_L) = R_F \parallel R(2 + g_{m,\text{eff}} R) \tag{44}$$

用 R_F 的表达式替换 R_F 后，上式变为

$$R(1 - A_V) \parallel R(2 + g_{m,\text{eff}} R) \tag{45}$$

注意，在增益很大的极限情况下，面向 C_{gd} 的电阻接近于：

$$|A_V|(R/2) \tag{46}$$

这从考虑米勒效应中就可以预见到。

面向 C_{gs} 电阻的计算更为繁琐些，但最终可以推导出如下的表达式：

$$\frac{R(R_F + R + 2R_E) + R_E R_F}{(2R + R_F)(1 + g_m R_E) + g_m R^2} \tag{47}$$

在增益很大的极限情况下，这个量可简化为

$$\frac{R}{R_E}\frac{1}{g_m} \tag{48}$$

注意，比值 R/R_E 近似等于增益（从基极至集电极）的幅度。由于两个开路电阻大致都正比于增益，所以并联-串联放大器的增益-带宽积近似为常数。

因此，在这种极限情况下，该放大器的带宽估计为

$$(\text{BW}) \approx \left[|A_V|\left(\frac{C_{gs}}{g_m} + \frac{RC_{gd}}{2}\right)\right]^{-1} \tag{49}$$

在推导了带宽的近似表达式后，我们现在来考虑输入阻抗。正如前面已经提到的那样，由于 C_{gs} 和米勒效应放大了的 C_{gd}，输入阻抗应当具有电容成分。通过假设在栅端的阻抗控

制了放大器的带宽，就可以得到总电容的粗略近似。也就是假设放大器极点的时间常数是电源内阻 $R_S(=R)$ 和在该节点上电容的乘积。基于这个假设，等效的输入电容就是式(49)中方括号内的项除以 R：

$$C_{in} \approx \frac{C_{gs}}{g_m R_E} + C_{gd}\frac{|A_V|}{2} \tag{50}$$

在几乎所有实际的情形中，由米勒效应增强的 C_{gd} 都占支配地位。

注意，由于这个出现在栅和地之间的电容的存在，实际上不可能做到在所有的频率下都实现完美的输入阻抗匹配。而且随着频率的提高，C_{gs} 逐渐变为短路，从而把源极的简并电阻 R_E 连到了栅极节点上。因此甚至输入阻抗的电阻部分也在下降，并随频率的升高而逐渐消失(译者注，这个说法不一定正确，是阻抗随频率增加最终消失)。

这些影响可以通过一些简单的方法在一定程度上得以减轻。首先，可以采用一个 L 匹配使电阻部分在某个标称频率处(这个频率一般比匹配质量开始显著下降时的频率略微大一些)得以提升，例如至 50Ω。在可能的 L 匹配类型中，最好的选择通常是把一个电感与栅串联并且在放大器的输入端并联一个电容，这样一个网络在低频时不起什么作用，因为那时不需要有任何校正。

L 匹配中的串联电感通常会留下残余电感成分。加大 L 网络中的并联电容可以很容易地补偿这个电感。进行这种补偿后，能达到良好的输入匹配的频率范围常常可以加倍。

为了计算输出阻抗，将一个测试电压加到集电极并计算测试电压与它产生的电流的比。在增益很高的极限情况下，我们发现输出阻抗包含一个电感成分，其值近似为

$$L_{out} \approx ARC_{gs}/g_m \tag{51}$$

式中，C_{gd} 已被忽略。

为了发展一个对这个电感来源的更深层次的理解，我们注意到基极电压是加在集电极测试电压的一部分。具体来说，由于基极电容的存在，基极电压是集电极电压经过衰减和低通滤波后的结果，因此基极电压滞后于集电极上的电压。进一步，晶体管把延迟的基极电压转变成延迟的集电极电流。从测试电源的角度来看，它一定提供了一个电流的一部分是滞后于所加的电压。这个电压和电流的相位关系就是电感的特征。

由以上的内在理解分析，可以估算出因忽略 C_{gd} 所带来的影响。由于 C_{gd} 在基极提供了一个领先的电压成分，所以它往往可以抵消这个电感效应。因此，如果 C_{gd} 不是小到可以忽略不计，那么实际观察到的输出电感比式(51)所估计的上限要小得多。

设计实例

现在考虑一个设计例子，这个例子会加深我们对一些重要概念的理解并进一步迫使我们去采取措施来面临偏置提供电流的挑战。我们要设计一个通用的 10dB 增益模块，为了便于级联，其输入输出阻抗应为 50Ω。在这种情况下，我们寻求的是实现 3.2 倍的电压增益。根据一阶近似公式，我们需要一个发射极的简并电阻，大约为 50Ω×(1+3.2)=210Ω。但是一阶增益公式是一个乐观的估计，因为它忽略了晶体管自身的输出电阻。所以，先随意定一个值，选 R_E=10Ω，然后稍后再对其进行一些修改。

下一步，选择反馈电阻值

$$R_F \approx Z_0(1-A_V) = 50\Omega \times (1+3.2) = 210\Omega \tag{52}$$

将它设为 220Ω，这是我们能有的误差最小的标准电阻，其容差为 10%。就像 R_E 一样，最后

定的 R_F 将可能需要修改。

现在已经完成了放大器核心部分的设计，我们需要选择一个晶体管的偏置点并找出建立偏置的方法。其中一个考虑是我们希望来验证一下(至少是适度地验证)增益估算公式。也就是说需要选择一个足够大的集电极偏置电流以使 $1/g_m$ 加到 $R_E=10\Omega$ 不会比一阶的估算值 12Ω 高。如果选择 10mA 的 I_C，那么 $1/g_m$ 在室温下大约为 2.5Ω，这是一个合理的近似值。尽管它略微超出了限制，我们还是选用它了。注意到偏置电流也设定了输出摆幅范围(也就是输出功率)。取决于设计目标，偏置电流可能需要再增加一些。对于这个设计例子，首先假设输出摆幅的需求不是太高，因而 10mA 也就够了。

下一步，需要在这个集电极电流下选择一个具有可接受特性的晶体管(比如说，足够高的截止频率和小的寄生参数，等等)。在这个例子的下面部分，假设用一个 2SC3302 微波晶体管，它的特性是足够满足要求的。

现在我们已经选定了偏置条件，要设计一个网络以稳定地建立此偏置。假设选择的拓扑结构如图 12.28 所示。我们已经加入了几个元件去产生一个稳定的偏置，但是 AC 电路仍然没有改变。电阻 R_B 上的压降乘以 R_F 与 R_{B2} 的和产生集电极电压。为了验证此结论并确定对晶体管参数 β 最小的依赖，要求通过这些电阻的电流相对于基极电流要足够地大。同时也不希望对输入端的负载太重，否则增益和输入匹配都会受影响。为了量化可接受值的范围，假设正常基极电流为 100μA，以获得 10mA 的集电极电流，基极电压为 1V。这样 R_{B1} 必须小于 1kΩ 以使得通过它的电流至少是基极电流的 10 倍。为了避免过载，我们希望 R_{B1} 不小于 10 倍的信号源电阻，或 500Ω。我们任意选一个值 680Ω，因为它在这两个界限值中间，而且这个值对应于一个容差为 10% 的标准电阻。这个值对应于通过 R_{B1} 的电流大约是 1.2mA。

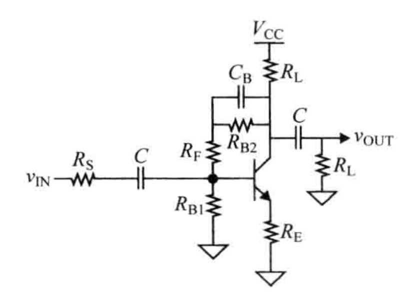


图 12.28 完整的并联-串联放大器

下一步，假设电源电压为 10V。因为结电容随着反向偏压的增大而减小，我们可能希望 10V 的电压尽量都落在了晶体管上。然而，我们需要留一部分压降于负载电阻上以确保一定的输出摆幅。如果没有更具体的设计指标，就随意选择降一半的电源电压在 R_L 上。这样我们选定 R_L 等于 5V/10mA=500Ω。

为了完成偏置网络元件的选择，我们希望在 R_F 和 R_{B2} 上降 4V 的压降，这样如果 1.2mA 的电流流过偏置网络，那么总的电阻大约是 3.3kΩ。然后将 R_F 从这个和中减去得到大约 3kΩ。我们将其取为 3.3kΩ，这是距设计值误差为 10% 的最近的一个标准电阻器，在上面的演算中，我们实际上已经忽略了在 R_F 上的 DC 压降。

最后是耦合电容 C，只要在最低的工作频率处的阻抗比 50Ω 小得多即可。如果我们违反了适度的原则使用了太大的电容，那么它的寄生参数(比如串联电阻和电感)带来的不良效应在高频时将很大。如果放大器的工作范围比较宽，我们可能需要并联几个电容。并联组合是比较合理的，因为这样能保持信号路径的物理宽度恒定并避免引入阻抗沿着信号路径发生不连续的现象。

和往常的一样，我们也需要多少考虑反馈旁路电容 C_B 带来的寄生效应。幸运的是，正如我们将要看到的，并不需要太高的性能。对眼下的应用，只需选择一个电抗值至少 10 倍

小于 R_{B2} 的电容即可。

图 12.29 所示的仿真曲线显示这个初步的设计已经很接近设计目标。偏置电流是 9.5mA，增益将近 9 个 dB。前者已经很接近目标因此不需再做调整。如果愿意的话，在 1dB 增益存在的偏差可以通过减小射极简并电阻值几个欧姆来消除。

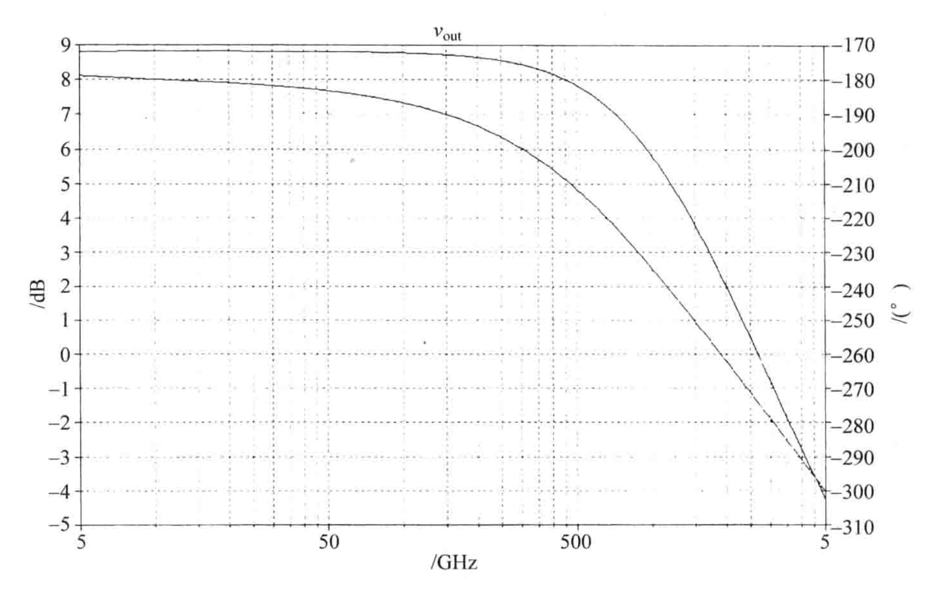


图 12.29 第一轮设计的 LTSpice 模拟结果

我们并没有给出这次设计的带宽要求，但是注意到仿真揭示了带宽在 1GHz 左右。巧合的是，这个值正好与研究带宽扩展技术中的未经补偿的开环放大器的一样。因此如果需要更大的带宽，可以采用相同的带宽扩展技术。其中最直接的方法就是在并联反馈途径中加入一个串联电感来创造一个零点，正如并联成峰放大器那样。如果这个零点放置得恰当，它可以补偿主极点，将频带拓宽。在这个例子里，我们想选择一个在 1GHz 附近的零点。因为反馈电阻是 220Ω，我们就将电感的初始值选为

$$L \approx \frac{R_F}{\omega_{zero}} = \frac{220}{2\pi \times 10^9} \approx 35(\text{nH}) \tag{53}$$

图 12.30 的仿真显示了加入电感后的结果。正如所看到的，带宽现在大约是 1.7GHz。注意，提升的倍数和并联成峰技术提供的相似。0.5dB 的增益上翘不是那么地不可接受，但是必要的话，可以通过减小反馈电感值来加以消除。

因为反馈路径里有电感能帮助提升带宽，我们不需要过分担心电容 C_B 的电感寄生成分。同时因为反馈电阻值本身就很大，我们也不需要过度担心电阻寄生成分。但是，我们总是需要验证这些假设。

第一轮设计显然是不完整的，但是只需要简单易行的优化就能完成这个设计。我们把它留给读者作为练习。

最后，检查一下如何改进这种拓扑结构输入匹配。尽管并联-串联反馈拓扑结构在两个（即输入与输出，译者注）端口处能十分有效地提供宽带的匹配，但匹配在任意宽的频带范围内绝不会是完美的。例如，在输入端口，晶体管的容性的输入阻抗会随频率升高导致匹配

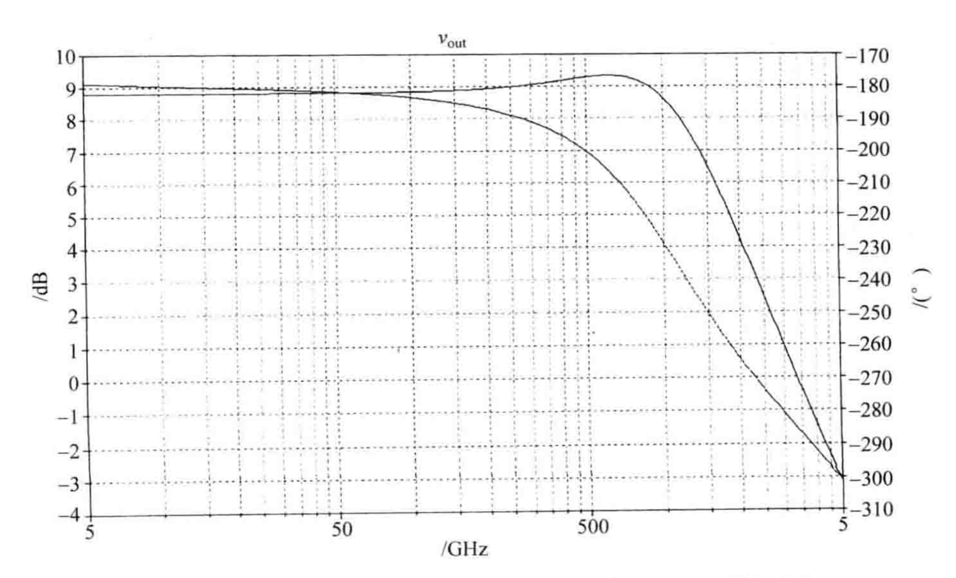


图 12.30　带有反馈电感的并联-串联放大器的 LTSpice 模拟结果

变差。我们也会发现实部在典型情况下也在减小。这些效应是在频带边缘处增益降低引起的自然结果。当增益降低时，反馈效果将减弱，因此输入匹配将接近开环状态。

如果希望拓展具有良好匹配的频率范围，需要重新设计并联-串联放大器（例如，以某种方式修改一下反馈网络）或者简单地在放大器前级联一个匹配网络（而不去改动放大器的拓扑）。如果选择后一种策略，必须限制自己在那些不改变低频下匹配的方法，因为低频下并联-串联放大器已经提供了很好的匹配。另外，寻求的匹配网络最好也不要求复杂的偏置。

作为一个具体的例子，考虑在 1GHz 带宽处对输入阻抗的测量告诉我们放大器在这个频率处的输入阻抗可以用 25Ω 电阻和 1pF 电容的并联来建模。有很多匹配网络可以让我们把它变换成 50Ω，但是并不是所有的匹配网络都符合我们这里已经确定的原则。作为一个不令人满意的例子，我们试图用一个适当的并联电感去谐振掉电容，然后将实部上变换。这种方法至少有两个缺点。一个是并联电感在直流下会造成对地的短路。解决这个问题可以通过串联一个大的电容，而这势必会引入寄生效应。另外一个更严重的问题是这个网络将彻底毁掉低频下的匹配。因此，我们虽然可以解决 1GHz 处的匹配问题，但同时毁掉了其他频率处的匹配。

一个更好的选择是使用简单的低通 L 匹配（见图 12.31）。串联电感 L_m 和并联电容 C_m 对偏置也有利。此外，这种网络在低频没有影响，我们因此可以保留低频下已经能提供的良好的匹配。

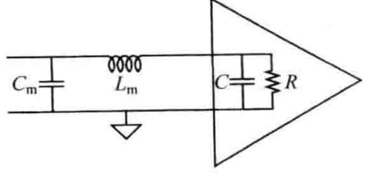


图 12.31　一个提供在高频下匹配改进的恰当方法

在史密斯圆图轨迹上，串联电感可以被选择来旋转阻抗直到在 $G=1$ 的圆上。然后，并联电容将其旋转到史密斯圆图的中心位置，这样就完成了匹配。取决于设计目标，我们可以在下面两个频率点之间选择一个频率来实现完美的匹配：(a)未补偿放大器表现出差的匹配的频

率，(b)补偿放大器表现出可接受的匹配的上限频率。当只在匹配带宽内寻求适度的改善时(比如说，百分之几十)，简单的迭代就足以收敛到可接受的设计。

匹配网络可以在低频下用集总元件实现，在高频下用微带线实现。一个代表性的版图如图 12.32 所示。如果需要一个宽带的匹配(比如，两倍左右的带宽)，可以采用第四章中更复杂的谐振网络。当寻求一个特别宽带的匹配改善时(比如，接近十倍或更高的频率比)，最好的策略是采用一个切比雪夫滤波器，它的最后一级就是放大器的输入阻抗。这样有可能将带宽改善到 Bode-Fano 极限内的相当大的一部分。

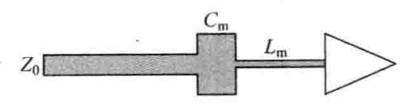


图 12.32　放大器匹配网络的版图

12.4.2　分布式放大器(行波放大器)

毫无疑问，分布概念的最巧妙的运用是英国的 William S. Percival[⑩] 在 1936 年发明的分布式放大器。然而，关于这个放大器他本人似乎并没有谈及很多，直到 1948 年在 Ginzion、Hewlett、Jasbefg 和 Noe 发表了一篇里程碑式的论文之后这一技术才被广泛地了解[⑪]。

在他们论文的摘要中作者评论道，"通常，'最大带宽-增益积'的概念并不适用于这个分布放大器。"让我们来看一看这一结构如何实现增益与延时的互换而不影响带宽。

如在图 12.33 中可以明显看到的那样，晶体管的输入是由一条带抽头的延迟线来提供的；而晶体管的输出则送入到另一条带抽头的延迟线。尽管这里显示的是简单的线段，但正如前面已讨论的那样，最佳性能可以在采用 m-导出或 T 形线圈线段的时候获得。分布式放大器因此是采用复杂网络实现增益-时延-带宽折中的一个逻辑的继续发展。

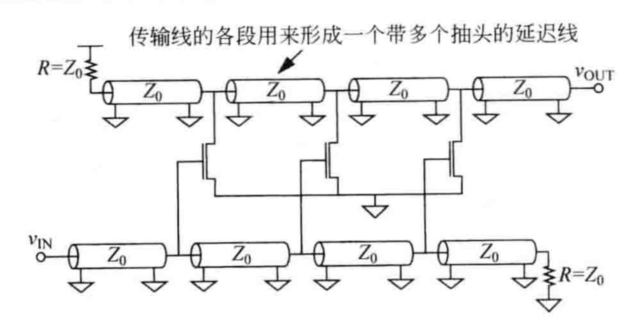


图 12.33　分布式放大器(场效应管版本)

加在输入端的一个电压阶跃将沿输入线向下传播，使这个阶跃依次出现在每一个晶体管上，每个管子产生的电流等于其 g_m 乘以输入阶跃的值。如果输入线和输出线的延迟相匹配，那么所有晶体管的电流最终将按照时间相干的方式被加在一起。

由于输出线上的每一个抽头都代表了一个阻抗 $Z_0/2$，所以总的电压增益(忽略损耗)为

⑩　W. S. Percival. Thermionic Valve Circuits(热离子阀电路). 英国专利号 460562，1936 年 7 月 24 日登记备案，1937 年 1 月授权。

⑪　E. L. Ginzton, W. R. Hewlett, J. H. Jasberg 和 J. D. Noe。Distributed Amplication(分布式放大器). *Proc. IRE*, pp. 956-969，1948 年 8 月。

$$A_V = ng_m Z_0/2 \tag{54}$$

与通常的级联放大器不同,我们看到这一放大器的理想增益与电路的级数成线性关系。注意到在这种连接方式下,放大器结构在本质上是一种相加的结构,放大器的总增益是各级增益之和,而非各级增益之积。因而放大器整体可以提供超过 1 的增益,即使放大器中的某级在某些频率下增益并没有超过 1。传统的级联放大器并不能实现这种奇迹。因此分布式放大器可以比通常的放大器工作在明显更高的频率下,而且由于延时也正比于级数,所以这一放大器确实以延时换取增益;而带宽确实没有以任何直接明显的方式影响这一交换。实际上,带宽的限制主要来源于随频率增加而增大的线上衰减。而这种线上的衰减部分又是因为晶体管的有损寄生参数导致的。

对这一放大器优势的另一种理解是我们注意到在传统放大器中限制带宽的一个原因是:当频率提高时,由于输入电容而使输入阻抗下降。与随频率增大的损耗性衰减不同,集总放大器的带宽限制主要来源于随频率增加而增加的反射。行波放大器通过将器件的输入电容吸收到了抽头延迟线的常数中而规避了这个问题[12]。因此通过利用这个电容来产生延迟,放大器的输入阻抗保持不变且等于 Z_0,直到接近传输线本身的截止频率。设计这种放大器重要的一点就是采用足够的级联单元以确保带宽不受延迟线的截止频率限制。尽管在这个例子中传输线是集总和分布元件的混合体,但基于集总元件假设还是可以给出关于截止频率的最坏情况的估计(有时被称作 Bragg(布拉格)截断)。也就是,

$$\omega_c \approx 2/\sqrt{LC} \tag{55}$$

这里电感是晶体管之间传输线的总电感,电容是线和器件本身的总的电容。我们总是可以将线分成足够小的段以将截止频率提高。

几个因素会共同导致行波放大器的功能变得不那么完美。其中最重要的因素通过考虑一个晶体管的简单模型(在这个例子里,是一个 MESFET,但它也可以是一个 MOSFET 的模型,只需调整一些参数)可以得到认识,见图 12.34。

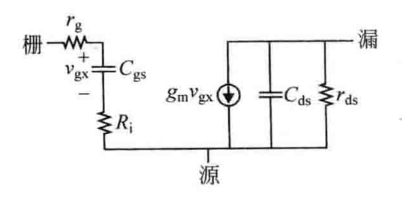


图 12.34　简单的增量 FET 模型

注意到这个模型忽略了 C_{gd},因为它的存在会阻碍撮取有用的设计直觉。因为同样的原因,我们也忽略了 3 个端口的电感性串联寄生参数。我们同时忽略了 NQS(非准稳态)效应引入的栅极-源极的并联电导。这个忽略可能对分布式放大器来说会是很严重的,因为行波放大器的带宽非常宽以至于在其高频工作区域 NQS 的负载效应已经十分明显了。

重要的是要注意到由这个模型会得到晶体管对输入端和输出端的负载影响不同。输入端的电容是大于输出端的电容的,所以需要作一些大的调整以保证匹配的延时和阻抗。满足这些要求需要漏极线比栅极线更长和更宽以使延时和特征阻抗相等。我们必须满足的一些专门的条件大致如下:

$$Z_{0,in} \approx \sqrt{L_{in}/(C_{in} + C_{gs})} \tag{56}$$

[12] 在这里,我们假设器件的输入阻抗在高频时可看作容性。然而,这个假设并不总能满足,在实际设计中,为了达到更好的性能,需要考虑偏离这个假设的情形。

$$Z_{0,\text{out}} \approx \sqrt{L_{\text{out}}/(C_{\text{out}}+C_{\text{ds}})} \tag{57}$$

$$T_{\text{d,in}} \approx \sqrt{L_{\text{in}}(C_{\text{in}}+C_{\text{gs}})}(l_{\text{in}}) \tag{58}$$

$$T_{\text{d,out}} \approx \sqrt{L_{\text{out}}(C_{\text{out}}+C_{\text{ds}})}(l_{\text{out}}) \tag{59}$$

在这些等式中，L_{in}和C_{in}代表抽头间线的总的电感和电容（我们假设线段的长度比其波长远小，这样可以把它当作集总元件）。这些抽头间的延时为$T_{\text{d,in}}$。相应的输出线上的那些量用下标 out 表示。一般情况下，我们希望能有相同的阻抗和延时。

从以上等式可以看出，没有加晶体管的输入输出传输线的特征阻抗必须被设计成大于Z_0。此外，因为晶体管输入电容对输入线的负载效应强于其输出电容对输出线的负载效应，设计的输入线的特征阻抗要大于输出线的特征阻抗。就像在多数情况下，如果我们的设计目标是输入输出端口具有相同的特征阻抗Z_0，那么我们可以推导出关于传输线常量的一个特定的条件如下：

$$Z_0 \approx \sqrt{\frac{L_{\text{in}}}{C_{\text{in}}+C_{\text{gs}}}} \Rightarrow T_{\text{d,in}} \approx \frac{L_{\text{in}}}{Z_0}(l_{\text{in}}) \tag{60}$$

$$Z_0 \approx \sqrt{\frac{L_{\text{out}}}{C_{\text{out}}+C_{\text{ds}}}} \Rightarrow T_{\text{d,out}} \approx \frac{L_{\text{out}}}{Z_0}(l_{\text{out}}) \tag{61}$$

令输入输出时延相等得到

$$(L_{\text{in}})(l_{\text{in}}) = (L_{\text{out}})(l_{\text{out}}) \Rightarrow \frac{L_{\text{in}}}{L_{\text{out}}} = \frac{l_{\text{out}}}{l_{\text{in}}} \tag{62}$$

因此，选择的线段的长度与其电感成反比。输入线段必须要有比输出线段更高的电感值来平衡阻抗，而输出线要长于输入线以平衡时延。

目前我们仍然缺少足够的信息来进行一个实际的设计，要这样做，需要给定的设计指标和约束条件。比如：可以寻求最大增益而无其他约束；在固定的线段长下获得最大增益；固定增益下寻求带宽最大化；或者在给定功耗下获得最大的增益带宽积。不同的设计需求导致不同的设计。我们这里只提供如何实施这些策略来达到这些指标中某些的大纲。

对于给定的晶体管和固定的线段长度，如果在没有功耗限制的情况下希望增益最大化，那么优化方法会给出对应于这些细致的电路级数。在我们增加级数以提高增益的同时，总的线长也在增加。由于线长增加会引入指数式的衰减，最终会达到这样一个情况：增加一级（放大电路）引入的衰减反而比增益多。能获得最大增益的准确级数是关于器件的工艺和尺寸、偏置点，以及传输线常数的函数。对于这个学术上的条件，可以证明最优的级数大约为

$$N_{\text{opt}} \approx \frac{\ln(\alpha_{\text{in}} l_{\text{in}}/\alpha_{\text{out}} l_{\text{out}})}{\alpha_{\text{in}} l_{\text{in}} - \alpha_{\text{out}} l_{\text{out}}} \tag{63}$$

其中，α_{in}，α_{out}分别是输入输出线路的衰减常数（考虑晶体管的负载效应在内）[13]。注意到上式并没有给出如何选取最优的晶体管和偏置的向导。相反，假定晶体管规格和偏置是给定的（由此可以计算晶体管对线负载的贡献）。在集成电路设计中，我们能完全控制器件的宽度。即使在分立元件设计条件下，我们至少也能选择器件的某些尺寸，但是上述优化方法却没有利用那个设计自由度。同时注意到，此设计方法并没有涵盖达到特定的带宽指标。

作为一个更实际的（但复杂得多的）选择，我们通常给定晶体管的偏置电压和总的栅宽。

---

⑬ 见 David M. Pozar, *Microwave Engineering*（微波工程），2nd ed. Wiley, New York, 1998.

这些约束条件实际上确定了总的功耗和增益。因此，在这种情况下，最优级数的确定就产生了一个设计，其带宽在固定的功耗预算下取得最大化。

在寻求任何一个上面的优化目标时，最重要的是要认识到电阻性器件寄生效应会在改变输入输出线路的特性上带来很大的麻烦。比如，考虑到非准静态晶体管动力学（图 12.34 中忽略了）就出现了一个随频率的平方而增大的并联电导。这个 NQS 损耗是叠加到其他的器件损耗，以及在所有实际传输线路内部存在的再损耗。传输线的损耗和频率相关，这使得问题更加复杂化。若趋肤效应占主导，则损耗随频率的平方根增加；若电介质损耗占主导，则损耗随频率线性增加；若两种机制都需要考虑，损耗也就是前述两种表现的混合。

无论损耗因何而起，它都会带来衰减和色散。同时可以从晶体管的模型中看到，输入输出线通过栅漏电容而耦合。所有这些二阶效应都会降低电路性能，并给电路设计带来更大的挑战[14]。总的来说，我们需要结合轮流使用电路仿真器和电磁场解法器的迭代来最后得到一个好的设计。这些挑战并非是不可逾越的，这可以从行波放大器（TWAs）展现的创纪录的增益-带宽积清楚看出。

为了印证上面的观察结论，或许知道真空管分布放大器曾在 Tektronix 公司的许多示波器中成功地使用了许多年会有帮助（他们首次采用 513 型这类放大器）。这些放大器用在最后一级垂直偏转级中，典型的要用六七对“匹配”的真空管。大约 $\omega_T/2$ 的带宽容易实现，因此大约在 1960 年，100MHz 的通用示波器就生产了[15]。

尽管具有这些特性，我们也许有理由疑问，为什么这类放大器的使用并不那么普遍。一个重要的原因是它的功耗很大，因为需要许多级来提供要求的增益。另外一个原因是行波放大器的噪声相当大，这是由于晶体管的驱动网络一般是不满足噪声优化（即最小）条件的。还有一个原因是最初代替真空管的有源器件是双极型晶体管，它们有几个特性使得其不适合用在分布式放大器中。最大的问题是基极寄生电阻 r_b，它使延迟线的 Q 值变差，从而使延迟线的性能下降。因此，双极型分布式放大器得到了很不好的声誉。最后，集总参数的延迟线直到最近才能被集成，因为此时的器件已得到充分的改进，从而使工作频率提高到大全集成的延迟线已实际可行的频率范围，于是分布式放大器就几乎完全消失了。

分布式放大器最终又重新出现在 1980 年，当时研究砷化镓工艺的工作人员重新发现了这个原理。自此之后，分布式放大器采用各种化合物半导体工艺来实现。其中，采用 InP（磷化铟）的放大器达到了 100GHz 的带宽，如今有一种采用 0.18μm 工艺和键合线电感的 CMOS 实现达到了接近 5GHZ 的带宽。随着 CMOS 工艺技术的继续改进可望达到更明显高的带宽。

12.4.3　宽带放大器的交调失真

在有关低噪声放大器这一章（第 13 章）中，讨论窄带情况的重点，也扩展到表征非线性性的方法。三阶交调点作为线性性的一个度量的价值取决于它可以忽略系统频带外的谐波

[14] 当然我们总是可以使用 cascode 单元来减小耦合问题，这样做是以电源电压裕度降低为代价的。同时采用 cascode 结构，我们不得不设法兼顾因通过共栅器件的额外的延迟，这又增加了设计过程的复杂性。

[15] 在 Tektronix 585A 100MHz 示波器中的分布式放大器采用 6DJ8 双三极管，其 f_T 的值大约可以达到 300MHz。延迟线由 T 形线圈构成，这可以提供比普通的 m 推出的集总元件近似更好的带宽。而且在多数情况下，因为线圈之间十分靠近而提供了某些跨接电容，从而进一步提高了带宽。

和交调失真乘积的能力。

与此相对比，宽带放大器在一个很宽的频带内接受、放大也使得信号发生失真。因此，适用于窄带系统的失真测量方法可能不适用于宽带系统，比如说有线电视放大器和正在兴起的超宽带(UWB)系统。在这些例子里，三阶交调需要被另外的线性度测量手段所取代或至少是被加以补充。

有线电视可以作为一个实际的演示有代表性的例子。它有 91 个通道，每个 6MHz 宽，从首端传递到用户。因为总的带宽超过半个 GHz，所以很难滤除由于非线性产生的失真。除此之外，人眼对某些类型的信号假象格外敏感。因此，尽管信号-失真比数值上很大(例如，超过一般数字系统所要求的、完全可以忽略的误码率)，但这仍然无法确保接受信号的质量。因为有线电视系统在 19 世纪 70 年代大量发展起来，工程师逐渐意识到这些视觉上不可接受的失真。其中最重要的是复合三频拍击(composite triple beat，CTB)和复合二阶(composite second-order，CSO)失真。正如它们名字所隐指的，CTB 是由 3 个载波的交调产生，而 CSO 是由 2 个载波的交调产生。

考虑三个载波频率 f_1，f_2，f_3。如果它们挨得很近，只举一个例子，交调项($f_1+f_2-f_3$)会处于原始的载波频率附近。如果这个 3 调拍击乘积项不是足够小，它就会产生一个可见的"爬行"现象，除非来了大的载波-CTB 比信号，这个现象看起来很明显。注意，存在着其他的同样这 3 个频率产生的交调项，它们同样可以落在原始载波频率附近。实际上，任何涉及到一个和与一个差的组合都会产生一个在原始频率旁边的三阶交调项。

大量的实验揭示了当这些交调项比载波频率强度低 50dB 时，大多数人是很难辨别出来。因此 FCC 规定电视设备需要满足 51dB 的载波-CTB 比例。大多数商用的发送端的放大器一般会超过最低要求 10dB 或更多。

因为 CTB 是一个三阶的失真，理论上信号功率每降低 1dB，它会降低 3 个 dB。因此，对于功率下降 1dB，C-CTB 比例改善 2dB，而 SNR 只降低 1 个 dB。类似地，对于信号功率下降 1dB，C-CSO 改善 1dB。和 IP3 一样，这些理论上的推断在实际中未必被满足。我们总是可以构建一个放大器，它不符合这些关系。尽管这么说，这些粗略的规则总体上适用的时候比不适用的时候要多。

测试 CTB 类似于测试三阶交调。这一次是有三个信号产生器给器件提供信号。和测试 IP3 一样，信号产生器的输出应该相互隔离(或者在相加前加入各自的衰减器)以防止一个信号会扰乱另一个。

商用的仪器采用未调制的载波(CW)来进行 CTB 测试。这种条件要求是最苛刻的，因为调制的载波有着更小的平均功率所以会产生低电平的 CTB。典型的情况是，我们发现在主观的 CTB 测试中，采用实际的电视信号相比于 CW 测量会有 10dB 的改善。

12.5 调谐放大器

12.5.1 引言

我们已经看到，宽带放大器的设计可以通过并联成峰补偿一类的带宽扩展技术得到帮助，然而并不总是需要(或甚至不希望)在一个很大的频率范围内都提供增益。我们所需要

的一切常常就是在以某个高频为中心的窄频范围上的增益。

这样的调谐放大器广泛应用在通信电路中，它对所希望的信号提供有选择的放大而对不希望的信号提供一定程度的滤波。正如我们将很快看到的，消除了对宽带操作的要求使我们可以在比较高的频率上得到显著的增益。也就是说，对于零阶而言，为在 1MHz 的带宽上得到增益 100 所要求的努力大致上与所得到带宽的中心频率无关，在获得一个特定增益-带宽积时所遇到的困难几乎是恒定的而与中心频率无关(在一定限度内而言)。而且，对于一个窄带实现为得到这一增益所要求的功耗可以相当地少。这最后一个考虑在设计便携式设备时特别重要，因为此时电池的寿命是一个主要的考虑因素。

12.5.2　带单个调谐负载的共射放大器

为了理解为什么增益-带宽积会大致与中心频率无关，考虑图 12.35 所示的放大器。如果用一个零阻抗的电源来驱动(如图 12.35 所示)，并且如果可以忽略串联的基极电阻，那么集电极-基极电容 C_{gd} 可以归并到电容 C 中。在这种情形下，可以把这个电路建模成一个理想的跨导器驱动一个并联的 RLC 谐振槽路。在低频时，电感短路，所以增量(交流)增益为零，而在高频时，增益也为零，因为此时电容的作用如同短路。在谐振槽路的谐振频率处，增益则变为 $g_m R$，因为此时电感和电容互相抵消。

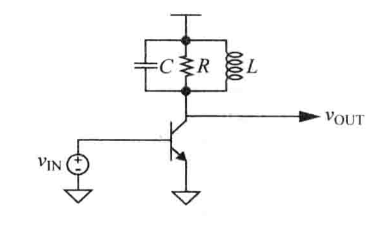


图 12.35　带有单调谐负载的放大器(具体的偏置被忽略)

对这个电路，总的 -3dB 带宽像通常一样就是 $1/RC$，因此增益(在谐振时测量)与带宽的积就是

$$G(BW) = g_m R \cdot \frac{1}{RC} = \frac{g_m}{C} \tag{64}$$

因此，对于这个例子，当满足所有简化的假设条件时，正如前面已指出的我们得到了不取决于中心频率的增益-带宽积。

为了强调上面这个论点的深刻含义，考虑两种不同的方法在 10.7MHz 时获得增益 1000(例如，对于调频收音机的中频部分)。可以尝试设计一个宽带放大器，它要求增益-带宽积超过 10GHz(这不是一个容易达到的目的)。或者我们可以意识到，对于调频收音机的例子，只需要在 200kHz 的带宽上达到这一增益[16]。在这一情形下，只需要达到差不多像 200MHz 这样的增益-带宽积，这是一个很容易完成的任务。

当然，在这两种方法之间的基本差别是在调谐放大器中负载电容为电感所抵消。只要我们能直接接触到任何寄生电容的两端(并且能使这些端出现在谐振槽路的两端)，那么就能选择合适的电感利用谐振来抵消它，从而在任意选定的中心频率处得到常数的增益-带宽积。

当然，实际电路并不会像这样所说的工作得那么好。比如，无论电感的质量是多么好，我们可能无法在 100THz 下从 Jell-O 牌(美国的一个食品商标，译者注)的果冻做成的晶体

[16] 这个值适用于商业 FM(调频)广播，你自己的应用可能会有所不同。

管中获得增益。但只要我们所选的中心频率是合理的[17]，那么调谐负载能使我们达到大致常数的增益-带宽积的这一点仍然是正确的。

12.5.3 调谐放大器的详细分析

上面进行的分析带有很多简化的假设。尤其是选择零电源内阻和零栅电阻使我们能把集电极-基极电容归并到谐振网络中去，从而使电感抵消它的影响。

因为假若我们无法直接将 C_{gd} 直接吸收到谐振槽路中去，这个电容可能有更为严重的影响，让我们考虑用比较实际的模型来模拟这个电路并考察一下会发生什么情况。具体来说，让我们现在用非零的电源内阻和非零的串联基极电阻，如图 12.36 所示。对应的增量(小信号)模型画在图 12.37 中。使用这个模型，可以计算两个重要的阻抗(确切地说是导纳)。首先，求出从 RLC 振荡回路向左边看过去的等效导纳，然后求出从电源内阻 R_S 向右边看进去的导纳。

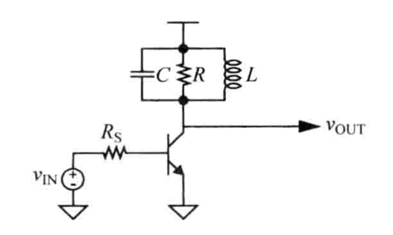


图 12.36　带有单个调谐负载的放大器

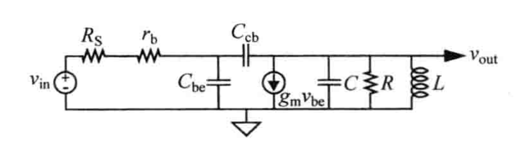


图 12.37　图 12.36 所示电路的增量模型

在进行这一分析时，在振荡回路的两侧加一个测试电压源以求出往它左边看到过去的等效导纳会好些。记住，无论你采用测试电压还是测试电流都将得到相同的结果(假定你没有在过程中出错，或至少你在两个推导时犯了同样的错误)，但这里采用测试电压比较方便，因为它允许最直接地确定了 V_{be}，而这个电压确定了受控电流源的值。

推导的完全过程是相当繁琐凌乱的，而且从这个过程中得到的回报也不大，但最终结果可以表述为：由振荡电路看到的导纳是一个等效电阻(我们现在先不去讨论它)与一个等效电容相并联。这个电容为

$$C_{eq} = C_{cb}[1 + g_m R_{eq}] = C_{cb}[1 + g_m(R_S + r_g)] \tag{65}$$

注意到 C_{eq} 会很大。实际上这是米勒效应的另一种表现，但现在是从输出口去看这个效应。加在集电极上电压的一部分出现在 v_{be} 上，它激励了 g_m 产生器。所产生的电流与通过电容的电流相加，由于它们都是由测试电源提供的，所以从测试电源看到了一个较低的阻抗。上述的电流的一个成分是来自简单的电容分压器，因而它与所加电压同相位，所以它代表了对振荡回路的电阻负载，这会引起增益下降。另一个电流成分领先于所加的电压，所以它代表了对振荡回路附加的电容负载。

C_{eq} 引起的附加电容负载效应使输出振荡回路的谐振频率降低，尽管这一偏移可以通过适度调整电感来补偿，但使电路工作在一个谐振频率，其在很大程度上取决于控制很差的、特性很差的以及有可能不稳定的晶体管寄生参数的情形，一般是不建议的。因此，与所预见

⑰　稍后，我们会更好地定量表示它，现在假设"合理"表示"合理地低于 ω_T"。

到的参数变化相比，我们希望把 C 选得较大一些，这样总的振荡回路的电容可以保持独立于工艺和工作点。但一个不幸运的结果是，对给定的跨导增益-带宽积是减小了。

在我们考虑输入阻抗（或者更直接地说是输入导纳）时，C_{cb}的更严重的影响就变得十分明显了。由于中间细节除了在推导我们将要说明的一小点结果之外同样没有什么用处，所以只列出推导的结果：

$$y_{in} = \frac{y_L y_F}{y_L + y_F} + \frac{g_m y_F}{y_L + y_F} \tag{66}$$

其中，已设 r_b 等于零；y_{in}是从电源内阻 R_S 往右边看到的导纳；y_F 是 C_{cb}的导纳，而 y_L 是 RLC 振荡回路的导纳[18]。

正如经常所发生的那样，如果反馈导纳 y_F 的数值与 y_L 的数值相比为较小时，可以写为

$$y_{in} \approx y_F + \frac{g_m(j\omega C_{cb})}{y_L} \tag{67}$$

在振荡回路显示为电感性的频率时（亦即在谐振频率以下）y_L 具有一个负的纯虚数部分，因此上式右边（因而也是 y_{in}）第二项可以具有一个负的实数部分。也就是说，这个电路的输入就好像有一个负的电阻与之相连。有负电阻就会引起振荡（如果这是你的意图那就正好，但一般的情况并不是这样），我们确实已有了振荡需要的所有元件——电感、电容和负电阻。如果这里没有 C_{cb}，那么本来是不会有这样的问题的。

因此，C_{cb}所带来的麻烦是它使输入和输出电路以产生坏的效应的方式耦合：它对输出振荡回路产生负载效应，并使增益减少，使输出振荡回路失谐，并可能引起不稳定性。若我们还希望把一个调谐电路加到输入端，那么这个问题将变得特别严重。而且，甚至在真正的不稳定发生之前，在两个端口调谐电路间的相互作用也可能使取得适当的调谐极其困难。

遗憾的是，C_{cb}总是不为零（事实上，它的典型值大约是主要的基极电容的 30%～50%，所以它很难被忽略），因此为了减轻它引起的各种不希望的影响，需要采用某些电路拓扑方面的技巧。

稍后我们会详细地探索放大器稳定性的问题，现在先考虑几个可以改善稳定性的技术。

12.6　中和与单向化

从了解到问题来自于输入和输出端口之间的耦合很自然就会想到一个策略，即移去这一耦合应当是有好处的。去掉输出和输入间的耦合应当是很熟悉的——它确实与在共发射极放大器中消除米勒效应一样。如果在那里能成功，则在这里一样能成功，见图 12.38。通过共基放大器在输入和输出端口之间进行隔离，我们可以消除（或至少大大抑制）失谐以及可能的不稳定性，从而可以得到较大的增益-带宽积。

达到这些目标的另一种拓扑连接是发射极耦合放大器（它可以看成是一个射极跟随器驱动一个共基放大级），如图 12.39 所示。同样，这一结构使输出与输入隔离，从而不会像在与简单的共发放大级中那样有严重的不稳定性和失谐问题。

---

[18] 为了避免进一步模糊已经不那么清楚的讨论，这里，已经忽略了晶体管的输出导纳；如果要进行更加准确的分析，它可以被吸收进 y_L 里面去。

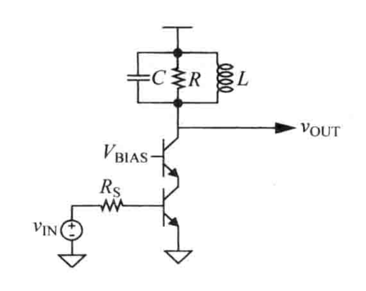


图 12.38 带有单个调谐负载的共发-共基放大器

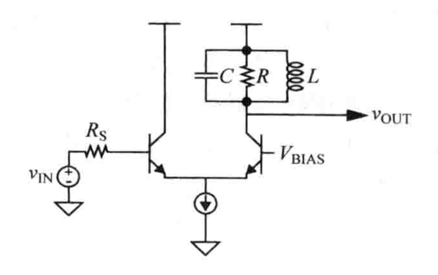


图 12.39 带有单个调谐负载的发射极耦合的放大器

就输入和输出隔离而言，共发共基放大器和发射极耦合放大器的作用是类似的。共发共基放大器对于给定的总电流可以提供大致两倍的增益(因为这一电流全部可以用来建立 g_m)，而发射极耦合放大器则对总的电源电压要求较低(因为两个晶体管并不像在共发共基放大器中那样上下堆叠在一起)。选择何种连接方式通常是基于如晶体管是否有足够的电源电压裕度以及增益应多大这样的考虑。

图 12.38 和图 12.39 所示的电路是接近"单向"放大器的例子，也就是在一个较大的带宽范围内信号只能沿一个方向流动。读者可以很清楚地看到这一单向性的价值。除了可以提供我们已经讨论过的电路的一些好处之外，它也使分析和设计电路容易得多.因为它减少或消除了不期待的和不希望的反馈。

如果我们不能(或选择不去)消除不希望的反馈，那么有另一种方法可以最大可能程度地去抵消它，由于这一抵消几乎不可能在较大的带宽范围内都很彻底，所以这一方法一般称为"中和法"[19]，以区别于能实现带宽更宽的不依靠抵消的宽带单向技术。

经典的中和放大器在图 12.40 中给出。注意，电感已用一个稍为复杂的元件所代替：一个抽头电感，它也称为自耦变压器。在所显示的连接中，由于对称性，这个电感上端和下端的电压恰好在相位上差 180°，[20]因此集电极电压与在平衡电容 CN 上端的电压间存在 180°的相位差。既然不希望的基极与集电极之间的耦合只是由于 C_{gd} 引起的，那么根据对称性，选择 C_N 等于 C_{gd} 就能保证从集电极至基极没有净的反馈。流过平衡电容的电流在数值上等于通过 C_{gd} 的电流，而在符号上与之相反，因此我们消除这个从输出至输入的耦合的方法是通过引入更多的从输出至输入的耦合来实现的(由于相位恰好相反，所以净耦合为零)。

中和法最初利用抽头变压器来实现，但片上变压器的品质因数较差(以及消耗较大的面积)使这一特定的方法对于集成电路(IC)实现并不诱人。然而，我们注意到抽头变压器只是用来得到反相信号，由于反相信号可以用其他方法很容易得到，所以更为实际的中和放大器仍然可以实现。一种电路拓扑是采用一个差分对来避免所需要的变压器，如图 12.41 所示。

[19] 中和法是在 1920 年代被当时为 Louis Hazeltine 工作的 Harold Wheeler 针对 AM(调幅)广播开发的。他发明的这种技术可以从 RF 放大器中获取大而稳定的增益，因此在典型的收音机中减少了所需的增益的级数(也因此减少了真空管的数量)，因而比许多对手的方法更加显著地减少了生产成本。

[20] 注意，自耦变压器在这里并不是一定需要的，它们只是在历史上用的比较普遍，而且当然是一个方便的手段来得到两个精确的相互相反的电压。显然，还有其他方法可以提供一对刚好相反的信号(考虑一下图 12.41 的例子)。

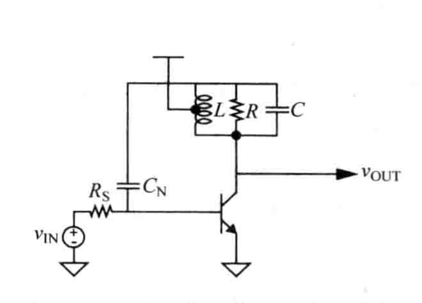


图 12.40　采用中和技术的共发放大器

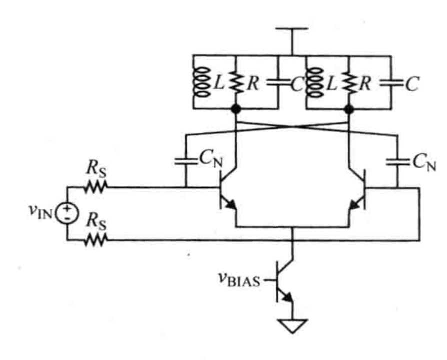


图 12.41　采用中和技术的共发放大器（对 IC 实现更为实际些）

因为采用这些技术能否达到完全的中和取决于能否送回一个电流与通过 C_{gd} 的电流完全相等，因此平衡电容 C_N 必须与 C_{gd} 完全匹配。遗憾的是，C_{gd} 的大小与电压有关。也许因为这一电容的变化使得完全的抵消不容易做到，所以中和法在半导体放大器中的应用很有限。而真空管由于它们的高线性度以及比较稳定的耦合电容，所以它们采用这一方法时情况会好得多。然而，如果足够努力的话，也可能在基于半导体的放大器中采用中和来获得有用的、大的增益-带宽改进。

12.7　异常的阻抗行为及稳定性

微波电路被那么多人认为是神秘的众多原因之一就是一些未预料到而出现的阻抗变换的数目。当然这些意想不到的现象只是因为你从来没有遇到过它们。我们现在研究一些这种奇异的阻抗行为(SIB)，主要的目的是使它们不再看起来奇怪。

为了演示一个通用的分析方法，同时将很多看起来没有联系的现象统一在一个上下文中，我们先来考虑为何从晶体管不同端口看进去的阻抗会取决于其他端口所连接的阻抗。我们将简化适用于 RF 领域的分析。更复杂的分析会导致更精确的答案，但是我们的目标是发展设计直觉，所以将限制与达到这个目标相一致的最小复杂度。

首先，考虑如图 12.42 所示的简单的晶体管的混合 π 形模型。这个模型左边的基极-发射极电压被当做独立的控制变量。这种选择和晶体管工作时实际的物理机理是接近的，但是并不是唯一可能的模型。从 12.42 的右图来看，我们也可以将基极电流作为控制变量。这两个方法是等价的，如果我们设置电流增益为

$$\beta = \frac{g_m v_{be}}{i_b} = \frac{g_m}{sC_{be}} = \frac{\omega_T}{j\omega} = -j\frac{\omega_T}{\omega} \tag{68}$$

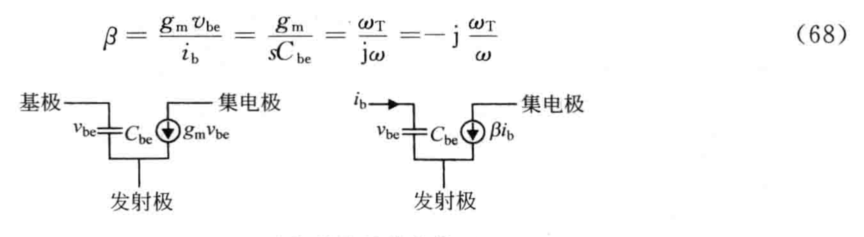


图 12.42　两种简单的晶体管模型

注意到等式(68)的意义：电流增益的幅度和频率成反比，最终在 ω_T 降到单位值 1。等式(68)也意味着电流增益有个正交的相位延迟，这是由输入电容上的电压正交相位延迟所导致的。最后，等式(68)表明，当频率接近零时，电流增益将增大到无穷大。这个错误意味着这个模型在电流增益远大于 DC 电流增益 β_0 的频率范围内已经不再适用，所以我们必须要小心地在适当的频率范围内使用这个模型（模型的低频限制对于 FET 来说，并不是那么大的问题，因为它有着很高的 DC 电流增益）。简单来说，电压控制模型预测在 DC 处有着无穷大的基极-发射极阻抗，而不是一个真正的有限的双极型器件的阻抗。再一次说明，只要在我们使用的频率处，输入电流主要是由输入电容所引起的，那么这个模型就可以产生合理的结果。最后，注意到这些模型可以适用于所有的晶体管，而不仅仅是双极型器件。因此，我们得出的结论可以广泛地应用于所有晶体管电路中。

在被已经认同与说明的限制下，这两个模型都可以产生相同的结果（它们是等效的），但是我们可以从一个或另一个中更容易获得设计直觉，但这取决于具体情况，所以有必要同时考虑这两种选项。

我们现在使用电流控制模型去推导发射极接任意负载 Z 时，向基极看进去的阻抗的表达式，见图 12.43。在这个例子中，更方便用测试电流源驱动基极。这样，降在发射极阻抗上的压降为

$$V_Z = Z(\beta+1)i_b \tag{69}$$

因此，从输入端口的观点来看，在那里只有 i_b 在流，发射极阻抗被乘了一个因子$(\beta+1)$，加上输入电容，我们得到总的输入阻抗为

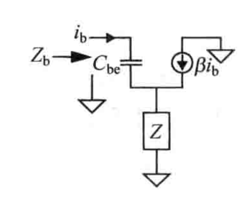


图 12.43 用于计算基极端阻抗的电路模型

$$Z_b = \frac{1}{j\omega C_{be}} + Z(\beta+1) = \frac{1}{j\omega C_{be}} + Z\left(-j\frac{\omega_T}{\omega}+1\right) \tag{70}$$

现在考虑一些特殊的例子，一个纯阻性的 $Z=R$ 变成一个电容（以及一个电阻），这个电容和基极-发射极电容串联，减小了总的输入电容。这种行为被认为是射级跟随器提供的基极-发射机电容自举的结果。电阻 R 的值越大，跟随器的增益越趋向于 1，导致的输入电容越小。

现在考虑发射极负载是一个纯电感，在这种情况下，

$$Z_b = \frac{1}{j\omega C_{be}} + j\omega L\left(-j\frac{\omega_T}{\omega}+1\right) = \frac{1}{j\omega C_{be}} + \omega_T L + j\omega L \tag{71}$$

也就是说，电感负载除了一个电感之外还贡献了一个纯电阻到输入端。给出一个参考数字，假设 1nH 的电感加入到发射极，如果晶体管的 ω_T 是一个不算大的 10GHz，那么可以在输入端口产生 63Ω 的输入电阻。显然，如果不想要的电阻性输入阻抗需要避免的话，减小发射极电路电感寄生参数是非常重要的。晶体管速度越快，这个挑战越大。然而，正如在低噪声放大器那个例子看到的，通过电感简并产生的输入电阻并非完全没有可用之处。一般来说，在某些情况下当一个效应足够大以至于产生麻烦时，就意味着在另一种应用下这可能是可以利用的效应。

现在考察电容性负载的效应，在这种情况下，

$$Z_b = \frac{1}{j\omega C_{be}} + \left(\frac{1}{j\omega C}\right)\left(-j\frac{\omega_T}{\omega}+1\right) = \frac{1}{j\omega C_{be}} - \frac{\omega_T}{\omega^2 C} + \frac{1}{j\omega C} \tag{72}$$

注意到输入阻抗是电容性的，串联了一个负电阻。如果负电阻的幅度超过了驱动基极的正电阻的幅度，将会导致不稳定。即使在真正不稳定行为发生作用之前，负电阻也会以产生谐振峰的方式改变输入阻抗并且降低输入匹配的质量。考虑到经常会遇到射极跟随器后接电容负载的情况，对这种机制的了解是至关重要的。

现在我们研究一下随基极阻抗的变化，由射极看进去的阻抗是如何变化的，见图 12.44。在发射极有负载的情况下，阻抗被倍乘了一个因子($\beta+1$)，这里，它是被除以同一个因子。那就是

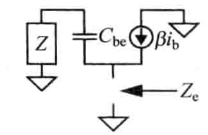


图 12.44　用于计算发射极端阻抗的电路模型

$$Z_e=\frac{1}{j\omega C_{be}}+\frac{Z}{\beta+1}=\frac{1}{j\omega C_{be}}+\frac{Z}{-j(\omega_T/\omega)+1} \tag{73}$$

为了将问题进一步简化，假设我们考虑频率足够低的情形，此时电流增益的幅度大于 1。在那种情况下，

$$Z_e=\frac{1}{j\omega C_{be}}+\frac{Z}{-j(\omega_T/\omega)+1}\approx\frac{1}{j\omega C_{be}}+jZ\left(\frac{\omega}{\omega_T}\right) \tag{74}$$

我们看到当从发射极看进去，基极串联的电阻变成了电感。这种看似由电阻创造出的电感有点让人迷惑，它是由基极-发射极电容和电阻之间作用产生的结果。事实上，电感的出现是由基极-发射极电容导致的回转效应产生的。再次指出，这种电感是有利还是有害要取决于电路的其他部分。取决于那些详细情况，好与坏的结果都有可能。

一个电容性的基极阻抗贡献一个电阻成分，而一个电感性的基极阻抗贡献一个负阻。正如电容性负载的发射极跟随器，这种负阻既可能有利也可能有害。工程师必须了解这两种情况以利用或抑制这种机制。继续这样做下去，我们可以算出当一个电感接在基极与地之间时由发射极看进去的阻抗为

$$Z_e\approx\frac{1}{j\omega C_{be}}+j\left(\frac{\omega}{\omega_T}\right)(j\omega L)=\frac{1}{j\omega C_{be}}-\left(\frac{\omega^2}{\omega_T}\right)L \tag{75}$$

所以阻抗的实部为

$$R_{in}\approx-\frac{\omega}{\omega_T}\mid Z_L\mid \tag{76}$$

这种电路轻而易举地产生负电阻的能力是它们为某些类型的振荡器所广泛采用的原因。然而，如果负阻是不希望的，这种情况显然强调了最小化栅极寄生电感的重要性。

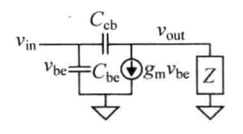


图 12.45　用于通用密勒效应分析的电路模型

我们在结束这一节的讨论之前，再重新看一下 12.4 节讨论的米勒效应，但这次略微有些不同。为了合理起见，将集电极-基极电容加入模型中，见图 12.45。而基极-发射极电容，因为它直接和输入端口并联，所以可以暂时将其移去，而在最终才考虑它。在输入端口加一个测试电压并计算输入阻抗(在没有 C_{be} 的情况)，得到

$$Z_{in}=\frac{1+sC_{cb}Z}{sC_{cb}(1+g_mZ)} \tag{77}$$

在 C_{cb} 的导纳比负载的导纳小得多的频率下，C_{cb} 对输入阻抗的贡献成分是

$$Z_{in}\approx\frac{1}{sC_{cb}(1+g_mZ)} \tag{78}$$

在集电极负载是一个纯电阻的特殊情况下，我们得的熟悉的米勒效应的容性倍增。

如果再假设跨导比负载的导纳大得多的话，那么

$$Z_{in} \approx \frac{1}{sC_{cb}g_m Z} = \frac{1}{j\omega C_{cb}g_m Z} \tag{79}$$

从式(79)可以看到纯电感性的负载为输入阻抗贡献了负电阻。这个结果和我们之前对调谐 RLC 共射放大器的分析是一致的。我们也可以看到个纯电容性的负载为输入阻抗贡献了纯电阻。

我们看到 R_F 电路里很多类型的阻抗变换，通常这些变换可以被利用，但同样经常的是不希望的。不论哪种情况，这里提出的方法使你有能力去做 0 阶的分析并让你找出它们的来源和一般的依赖性。这个知识通常足够用来找出补救方法，而不用去做更深入的分析，一直到设计阶段的最后，才需要更仔细的分析。例如，基极负阻的出现(比如，收集极有电感负载或者发射极有电容负载)可以通过加一些正的串联电阻。这些可以是简单的电阻或者在常见的分立 R_F 电路中采用铁氧体磁珠。后者的优点是在 DC 时没有电阻，因此可以最小地减小对偏置网络的影响。

这些阻抗变换经常迫使在放大器的接口处形成一些有意的失配。即使最大功率传输通常是很重要的，但在稳定操作的情况下也可能做不到。在输入、输出端口造成一些不匹配有利于保证稳定性。而且，在低噪声放大器的例子里，这种失配可能是必须的，因为这样可以改善噪声性能。

作为最后的评论，采用折中的设计往往是重要的。这里，设计一个放大器有过高的增益或过大的带宽都是不好的。更多未必更好。例如，源和漏端负载可能在带宽外性能退化。如果放大器的带宽非常宽，这些不受控制的阻抗可能导致不稳定。

稳定系数、稳定圆以及最大可获取增益

我们已经从器件级的角度研究了各种 SIB(奇怪的阻抗行为)的问题，以建立对它们产生来源、解决方法的基本理解。在经典的微波设计中，较常见的是以二端口参数和史密斯圆形式的稳定源、负载阻抗图的形式来讨论稳定性。尽管我们已经说过目的不在于重提经典的设计方法，不过关于这个话题还是值得说几句，至少可以使读者足够熟悉这样的数据，以便日后提供给他们时不至于感到陌生。

首先，遇到各种稳定系数的问题是很正常的。其中一个，由 Linvill 提出，是简单的判据，用来判断当输入输出端口都开路时，一个二端口网络是否会不稳定[21]。这个判据用器件作为二端口网络的导纳系数表示如下：

$$C = \frac{|y_{21}y_{12}|}{2g_{11}g_{22} - \mathrm{Re}[y_{21}y_{12}]} \tag{80}$$

要取得无条件稳定，我们需要 $C<1$。注意到 Linvill 稳定系数对反向跨导纳 y_{12} 非常敏感。用模型参数来表示，y_{12} 实际上就是 $j\omega C_{cb}$，所以选择小的集电极-基极(或者漏栅)电容是很有帮助的。我们在设计版图时也必须非常小心以避免不当地增加这两个节点之间的电容。另外，可以采用中和和单向化技术来减小 y_{12}，从而改善稳定性。

Linvill 系数的等式也告诉我们在输入输出端口加入电阻性负载可以改善稳定性。换

[21] J. Linvill 和 J. Gibbons. *Transistors and Active Circuits*《晶体管与有源电路》，McGraw-Hill，New York，1961。

句话说，它表述了这样一个直觉上能令人满意的观察结果：如果我们降低环路增益，那么稳定性就会提高。

因为 Linvill 测试的二端口网络参数假设端口都是开路的，所以 C 参数不能评估通常负载情况下的微波电路的稳定性。另一个稳定性系数是来自 Stern，允许我们去评估什么样的源、负载阻抗组合会带来不稳定[22]。由于采用了多少不恰当的分子分母表达式的随意选择，Stern 系数要大于 1 才表示稳定。Stern 系数定义如下：

$$K = \frac{2(g_{11} + G_S)(g_{22} + G_L)}{|y_{21}y_{12}| + \text{Re}[y_{21}y_{12}]} \tag{81}$$

式中，G_S 和 G_L 分别是源导纳和负载导纳的电导成分。和 C 系数一样，Stern 系数告诉我们当反馈电容减小（通过器件的选择或中和）或者在输入输出端口增加负载电导时稳定性会改善。

这两个稳定系数并不是唯一被定义或使用的，但是它们是迄今为止最经常遇到的。

在很多微波晶体管的数据手册里，制造商会善意地提供关于稳定性的信息，它会把处于不稳定区对应的阻抗区域画出来。这个信息让工程师知道不稳定性是否可能避免。作为一个简单的例子，考虑图 12.46 的演示。这个晶体管在右上部分的小圆范围内的阻抗下会表现出不稳定。有一部分阻抗是在单位圆外的，表示是负实部。在通常条件下，我们不会特意这样驱动一个放大器，但是正如当我们讨论 SIBs 时看到的，有可能无意间产生这样的负电阻。

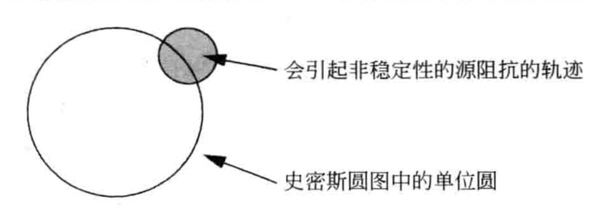


图 12.46　稳定性圆做法的图示

即使我们成功地避免了负实部，图 12.46 表示仍然有一些具有正实部的源阻抗会产生不稳定。这些是必须要避免的，但是这样做可能会妨碍其他一些目标的实现，比如最大化增益或是最小化噪声系数。

为了全面地评估所提出设计的稳定性，需要结合源稳定性圆来考察输出阻抗稳定性圆。

因为我们已经提到了增益，或许这里是时候来介绍一个品质因数，称为晶体管（或者任意二端口网络）的最大可取得增益。正如它的名字所指的，这个功率增益是在源共轭匹配和负载匹配同时取得时，并且器件已经单向化了的条件下可实现的，这时任何降低增益的反向跨导纳都等于零。

通过考虑一个包含输入和输出导纳以及一个跨导的二端口网络，我们可以为最大可取得功率增益推导一个简单的表达式。共轭匹配的条件意味着只有电导被留下。因此传送到匹配了的输出负载的功率为

$$P_{\text{out}} = V_{\text{in}}^2 \,|[\text{volatage gain}]|^2 g_{22} = V_{\text{in}}^2 \frac{|y_{21}|^2}{(2g_{22})^2} g_{22} = V_{\text{in}}^2 \frac{|y_{21}|^2}{4g_{22}} \tag{82}$$

㉒ Arthur P. Stern. Stability and Power Gain of Tuned Transistor Amplifiers 调谐晶体管放大器的稳定性和功率增益. *Proc. IRE*，1957，3。

而传送到二端口网络输入端的功率为

$$P_{\text{in}} = V_{\text{in}}^2 g_{11} \tag{83}$$

这样功率增益为

$$\text{MAG} = \frac{|\,y_{21}\,|^2}{4g_{11}g_{22}} \tag{84}$$

如果晶体管在所有负载条件下都是无条件稳定的，那么获得接近 MAG 的增益是有可能的。然而，在更实际的情况下，对应于最大增益的阻抗往往是不稳定的。类似的情况也发生在最小噪声匹配。没有人能保证一个器件可以同时实现最大增益、最小噪声和稳定的工作。

最后，我们重申一下采用适度设计思想的好处。有一种思维倾向是越多越好。但是选择高频性能远超过所需要的晶体管可能会产生严重的问题。晶体管提供增益的频率范围越宽，你就必须保证拥有稳定源和负载阻抗的频率范围越宽。经常会发生便宜且低性能的晶体管却可以实现更可靠的设计的情况。

12.8　附录　桥接 T 形线圈转移函数的推导

据作者所知，没有发表过关于桥式 T 形线圈(T-Coil)传输函数的推导，尽管最终的设计公式可以在一些文献中看到。遗憾的是，并不是所有这些公式的来源结论是一致的(尤其是对这里或那里出现的负号，不相同)，而且几乎没有哪个文献花时间来描述推导中所做的潜在的假设。因此，我们在这个附录里给出详尽的推导，保留尽可能多的中间步骤。

总的来说推导传输函数是直截了当但枯燥、繁琐的。为了减小在这个特定情形下推导过程中所要承受的痛苦经验，同时也可以借此减小出错的几率，我们采用第 7 章用来分析各种结构的偶模、奇模分解的方法。

我们将要分析的网络如图 12.47 所示。注意我们将单个跨接电容分解为两个串联连接的电容。这样做的原因马上就见分晓。

图中的电路几乎是对称的。不幸的是几乎的意思等同于不是，你可以得出结论，除了采用教科书里最原始的写出并求解基尔霍夫电压、电流公式外没有其他选择。幸运的是，这个网络是足够简单的，以至于这个电路最终还是可以采用对称的方法来解，尽管不那么容易。当然，这倒是一个可以作为惩罚学生的家庭作业，如果这些学生是要给予教训的。

我们不能希望能避免所有需要解的等式(毕竟这是一个 4 阶网络)，但是至少可以把问题分解为几小块，而每一块可以很直接地被解(及被检查)，因此大大地增加了得到正确结果的可能性。第一步是将问题改为对称的形式，如图 12.48 所示。

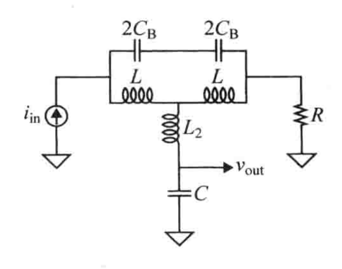


图 12.47　桥接 T 形线圈：原始的电路

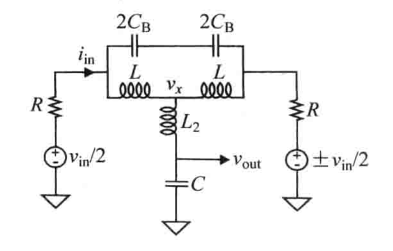


图 12.48　桥接 T 形线圈：采用共模与差模信号源的等效电路

现在我们已经将网络变换为完全对称的形式。这样就有可能采用差模-共模的分析方法来解决这个问题。尽管我们在输入端插入了电阻以确保能分解,但这实际上没有改变什么,因为我们只关心输出电压对输入电流的比例。“安培就是安培”(安培即指的是电流),至于输入电流是如何产生的,这一点并没有关系的。这种无关性同样允许我们放置一对电压源与电阻串联以分别评估差模和共模响应。叠加法将产生我们寻求的传输函数。注意到这种叠加原理最终会导致在右边的电阻下端产生一个零的净电压,这就对应于原来网络中的接地点。

12.8.1　差模响应

为了求得奇模传输函数,这两个电压源是有相同的幅度和相反的符号。多亏这种对称电路的非对称激励,我们可以将跨接电容和主电感的中间点接地,如图 12.49 所示。这种中间点接地意味着

$$v_{\text{out,d}} = 0 \tag{85}$$

这里下标 d 显式地表示变量是对应于差模响应的。

容易计算输入电流为

$$i_{\text{in,d}} = \frac{v_{\text{in}}/2}{R + [sL \parallel (1/2sC_{\text{B}})]} = \frac{v_{\text{in}}}{2}\left[\frac{2s^2LC_{\text{B}} + 1}{2s^2RLC_{\text{B}} + sL + R}\right]$$
$$= \frac{v_{\text{in}}}{2R}\left[\frac{2s^2LC_{\text{B}} + 1}{2s^2LC_{\text{B}} + s(L/R) + 1}\right] \tag{86}$$

图 12.49　差模分析的等效电路

注意,这个差模输入阻抗里含有两个纯虚数极点和一对复数零点。它们的自然频率是一样的,所以这个输入阻抗在 DC 和无限大频率处将趋向于同一个渐近值:$2R$。

12.8.2　共模响应

对于共模激励,跨接电容上没有压降。因为没有电流流过它,我们可以将它去掉,如图 12.50 所示。剩下的电路还可以用两种方式继续简化:一种可以将镜像对称的节点接在一起(导致 L 和 R 变为 $L/2$ 和 $R/2$,输入电流是我们要求电流的两倍);另一种方法是将网络在中间切开(这需要将 L 翻倍和 C 折半,这样输入电流就是我们想要求的)。这两种选择可以产生相同的结果,但这种选择可以基于其他考虑,比如方便性。在这里,哪一种结构都没有明显的优势,所以我们随意选择前一种,见图 12.51。

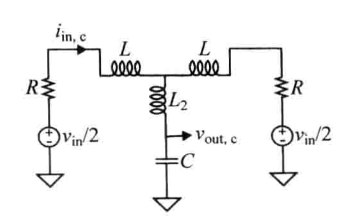


图 12.50　共模分析的等效电路

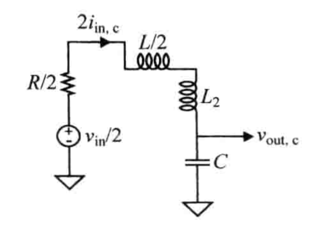


图 12.51　经简化后的用于共模分析的等效电路

输入-输出的电压传递函数就是一个简单的二阶 LC 低通滤波器：

$$v_{out,c}=\frac{v_{in}}{2}\left[\frac{1/(sC)}{s\left(\frac{L}{2}+L_2\right)+\frac{R}{2}+\frac{1}{sC}}\right]=\frac{v_{in}}{2}\left[\frac{1}{s^2\left(\frac{L}{2}+L_2\right)C+\frac{sRC}{2}+1}\right] \tag{87}$$

类似地，输入电流(这是我们要的电流的两倍)是输出电压乘以输出电容的导纳：

$$2i_{in,c}=(sC)v_{out,c}=\frac{v_{in}}{2}\left[\frac{sC}{s^2\left(\frac{L}{2}+L_2\right)C+\frac{sRC}{2}+1}\right] \tag{88}$$

或者

$$i_{in,c}=\frac{v_{in}}{2}\left[\frac{sC}{s^2(L+2L_2)C+sRC+2}\right] \tag{89}$$

12.8.3 完整的传输函数

在得到了偶模和奇模的输出电压和输入电流后，可以将这两部分的结果加起来得到输出电压和输入电流的比值。首先，总的输出电压为

$$v_{out}=v_{out,d}+v_{out,c}=0+\frac{v_{in}}{2}\left[\frac{1}{s^2\left(\frac{L}{2}+L_2\right)C+\frac{sRC}{2}+1}\right] \tag{90}$$

对应那个输出电压的输入电流为

$$\begin{aligned}i_{in}&=i_{in,d}+i_{in,c}\\&=\frac{v_{in}}{2R}\left[\frac{2s^2LC_B+1}{2s^2LC_B+s\frac{L}{R}+1}\right]+\frac{v_{in}}{2}\left[\frac{sC}{s^2(L+2L_2)C+sRC+2}\right]\end{aligned} \tag{91}$$

这个也可以写成

$$i_{in}=\frac{v_{in}}{2}\left[\left(\frac{1}{R}\right)\frac{2s^2LC_B+1}{2s^2LC_B+s\frac{L}{R}+1}+\frac{sC}{s^2(L+2L_2)C+sRC+2}\right] \tag{92}$$

因此，整个桥接-T 网络的跨阻为

$$\frac{v_{out}}{i_{in}}=\frac{\dfrac{1}{s^2\left(\frac{L}{2}+L_2\right)C+\frac{sRC}{2}+1}}{\left(\frac{1}{R}\right)\dfrac{2s^2LC_B+1}{2s^2LC_B+s\frac{L}{R}+1}+\dfrac{sC}{s^2(L+2L_2)C+sRC+2}} \tag{93}$$

下面我们一步步作简化，可以得到以下一系列的等式：

$$\frac{v_{out}}{i_{in}}=\left[\left(\frac{1}{R}\right)\frac{(2s^2LC_B+1)\left(s^2\left(\frac{L}{2}+L_2\right)C+\frac{sRC}{2}+1\right)}{2s^2LC_B+s\frac{L}{R}+1}+\frac{sC\left(s^2\left(\frac{L}{2}+L_2\right)C+\frac{sRC}{2}+1\right)}{s^2(L+2L_2)C+sRC+2}\right]^{-1} \tag{94}$$

和

$$\frac{v_{out}}{i_{in}}=\frac{2s^2LC_B+s\frac{L}{R}+1}{\left(\frac{1}{R}\right)\left\{(2s^2LC_B+1)\left[s^2\left(\frac{L}{2}+L_2\right)C+\frac{sRC}{2}+1\right]\right\}+\left(\frac{sC}{2}\right)\left(2s^2LC_B+s\frac{L}{R}+1\right)} \tag{95}$$

将分母中的各项相乘并合并同类项,我们最终得到了完整的 4 阶跨阻

$$\frac{v_{out}}{i_{in}} = [R]\frac{2s^2LC_B + s\dfrac{L}{R} + 1}{s^4L(L+2L_2)CC_B + 2s^3RLCC_B + s^2(LC + 2LC_B + L_2C) + sRC + 1} \tag{96}$$

观察到这里有 4 个极点和 2 个零点。通过目测发现,这两个零点是复数。有关极点,只通过观察我们说不出什么,只能指出如果它们是复数,那么它们会以共轭形式出现。

在开始实际的设计之前,我们需要加一些设计约束。

12.8.4 最大平坦幅度响应的设计公式

注意到取得最大幅度平坦度的一个必要条件是两个复数零点和一对复数极点相消,我们可以大大简化最大平坦响应的推导。然后就可以发现,那个剩下的极点对必须有最大幅度平坦度响应。不难证明后者在极点是复数,且有以下衰减因子时得以满足:

$$\xi = 1/\sqrt{2} \tag{97}$$

为了导出产生要求的极点-零点抵消结果的条件,采用综合相除法。执行完那个操作,衰减因子的要求就容易施加于剩下的二阶多项式上,这样就可以完成设计公式的推导。我们开始进行下面的运算,先从主要项开始(式(96)中的 R 只是一个比例因子,所以选择不将它写出)。

$$\frac{s^4L(L+2L_2)CC_B + 2s^3RLCC_B + s^2(LC + 2LC_B + L_2C) + sRC + 1}{2s^2LC_B + s\dfrac{L}{R} + 1} \tag{98}$$

在式(98)的第一步相除中,得到下面二次项

$$\frac{s^2}{2}(L+2L_2)C \tag{99}$$

和一个剩余三次项

$$s^3LC\left[2RC_B - \frac{L+2L_2}{2R}\right] + s^2L\left(2C_B + \frac{C}{2}\right) + sRC + 1 \tag{100}$$

再进行下一步相除得到线性项

$$\frac{s}{2}\frac{C}{C_B}\left[2RC_B - \frac{L+2L_2}{2R}\right] \tag{101}$$

及它相应的二次剩余项

$$s^2\left[2LC_B - \frac{LC}{2} + \frac{LC(L+2L_2)}{4R^2C_B}\right] + s\left[\frac{(L+2L_2)C}{4RC_B}\right] + 1 \tag{102}$$

最后一步则产生常数因子

$$1 - \frac{C}{4C_B} + \frac{(L+2L_2)C}{8R^2C_B^2} \tag{103}$$

和一个最后线性剩余函数

$$s\left[\frac{(2L+2L_2)C}{4RC_B} - \frac{L}{R} - \frac{LC(L+2L_2)}{8R^3C_B^2}\right] + \left[\frac{C}{4C_B} - \frac{(L+2L_2)C}{8R^2C_B^2}\right] \tag{104}$$

现在,在最终剩余函数的方括号里的部分必须等于零,因为只有这样才表明综合相除法能成功地产生二次因子,这个因子对应于没有被零点抵消的极点。那么,

$$\frac{C}{4C_B} - \frac{(L+2L_2)C}{8R^2C_B^2} = 0 \tag{105}$$

且

$$\frac{(2L+2L_2)C}{4RC_B}-\frac{L}{R}-\frac{LC(L+2L_2)}{8R^3C_B^2}=0 \tag{106}$$

解式(105)得到一个重要的关系

$$R^2C_B=\frac{L+2L_2}{2} \tag{107}$$

解式(106)得到另一个关系式

$$C_B=\frac{C}{4}\left[\frac{L+2L_2}{L}\right] \tag{108}$$

合并式(107)和式(108)最终得到一个实际元件的值

$$L=\frac{R^2C}{2} \tag{109}$$

下一步，注意到公式(103)的几个项相加为 1 这样，综合相除法发现了对应于未抵消的极点对的二阶多项式为

$$s^2\frac{(L+2L_2)C}{2}+s\left[RC-\frac{(L+2L_2)C}{4RC_B}\right]+1 \tag{110}$$

借助于式(107)和式(108)，可以将式(110)更加简化为

$$s^2R^2C_BC+\frac{sRC}{2}+1 \tag{111}$$

将式(111)表达为标准的二阶形式

$$\frac{s^2}{\omega_n^2}+\frac{2\zeta s}{\omega_n}+1 \tag{112}$$

即可得到

$$\omega_n=1/R\sqrt{C_BC} \tag{113}$$

$$\zeta=\frac{1}{4}\sqrt{C/C_B} \tag{114}$$

因此，还有另一个等式描述跨接电容和负载电容的关系。结合式(114)和式(108)，可以得到

$$C_B=\frac{C}{16\zeta^2}=\frac{C}{4}\left[\frac{L+2L_2}{L}\right] \tag{115}$$

然后可以通过联立式(109)和式(115)来推导出对余下这个元件 L_2 的等式：

$$L_2=\frac{R^2C}{4}\left(\frac{1}{4\zeta^2}-1\right) \tag{116}$$

注意，L_2 完全有可能取负值。实际上，我们看到桥式 T 线圈确实需要一个负电感因为有用的衰减因子超过 0.5。尽管孤立的无源电感无法提供负电感，但在变压器的模型中的互电感可以是正的，也可以是负的。幸运的是在我们的例子里，L_2 可以以这种方式精确产生。

到目前为止，我们唯一强加的条件是零极点抵消。因此，所有这些等式不仅是在最大平坦幅度响应的条件下适用，而在一般情况下也都通用。

总结一下设计步骤，假定给定 R 和 C。用式(109)计算主电感，用式(115)计算跨接电容以及用式(116)来计算剩下的电感，最后完成整个设计。注意进行设计比第一次推导这些公式容易得多了。

现在施加一个特定的条件——最大平坦幅度响应——来演示设计步骤。按照上面的步骤，首先找到主电感的值为

$$L = \frac{R^2 C}{2} \tag{117}$$

假定 $R=100\Omega$，$C=1.5\text{pF}$，找到电感 $L=7.5\text{nH}$。接下来，跨接电容可由式(115)容易得到

$$\left(\zeta = \frac{1}{4}\sqrt{C/C_B} = 1/\sqrt{2}\right) \Rightarrow C_B = C/8 \tag{118}$$

在我们的设计中它等于 0.1875pF。最终，

$$L_2 = \frac{R^2 C}{4}\left(\frac{1}{4\zeta^2} - 1\right) = -\frac{R^2 C}{8} \tag{119}$$

或者 -1.875nH，这样就完成了所有元件值的选择。

现在我们转向如何实现所需要的负电感。回想变压器可以建模成电感性的 T 网络。在两个相同电感耦合的特例里，模型如图 12.52 所示。其中电感 L_{pri} 是次级线圈(或初级线圈，对这个对称的情形)开路时初级线圈的电感。因此，这或许是在设计中需要注重的更为直接有用的量，因为这个是我们在为初/次级绕组画版图时需要确定的电感值。这两个绕组之间的耦合只会影响总电感在两部分间的划分，总和还是等于 L_{pri}。

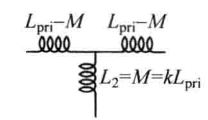


图 12.52　采用两个完全相同的耦合电感的等效电路

互感 M 的符号取决于两个主电感的磁场是相互加强还是减弱。如果是加强，那么互感是负值。比较图 12.52 的等效电路和我们的桥式 T 网络的元件，可以写出

$$L = L_{pri} - kL_{pri} = L_{pri}(1-k) \tag{120}$$

$$L_2 = kL_{pri} = \frac{kL}{1-k} \tag{121}$$

这样，可以将跨接电容表达式写成

$$C_B = \frac{C}{4}\left[\frac{L+2L_2}{L}\right] = \frac{C}{4}\left[\frac{L_{pri}(1-k)+2(kL_{pri})}{L_{pri}(1-k)}\right] = \frac{C}{4}\left[\frac{1+k}{1-k}\right] \tag{122}$$

相似地，主电感的公式可以改写为

$$\left[L = \frac{R^2 C}{2} = L_{pri}(1-k)\right] \Rightarrow L_{pri} = \frac{R^2 C}{2(1-k)} \tag{123}$$

现在，让人困惑的是，耦合系数经常被处理成是正的量，这样等式一般采用 k 的绝对值。在这种情况下，式(122)和式(123)分别为

$$C_B = \frac{C}{4}\left[\frac{1-|k|}{1+|k|}\right] \tag{124}$$

和

$$L_{pri} = \frac{R^2 C}{2(1+|k|)} \tag{125}$$

读者可以尝试去验证剩下的电感是负的，它的值为

$$L_2 = -|k| L_{pri} \tag{126}$$

在大部分发表的公式中，显式的绝对值符号是没有的，所以需要多加注意。至少有一个

例子里，符号的混乱导致一篇文章的作者声称给出了另一种设计方案据说可以避免使用耦合电感。除非存在二端的负阻电感，那篇文章的结论让人怀疑。

对于最大平坦幅度响应(巴特沃兹)的例子，耦合系数的值为 1/3。

最后，让我们通过计算在这些条件下的带宽，来估算带宽提升系数。首先，按照我们为最大平坦度计算得到的跨接电容，有

$$\omega_n = \frac{1}{R\sqrt{C_B C}} = \frac{1}{R\sqrt{(C/8)C}} = \frac{\sqrt{8}}{RC} \tag{127}$$

衰减因子依然为

$$\zeta = 1/\sqrt{2} \tag{128}$$

碰巧的是，这个特别的衰减因子对应了 -3dB 带宽精确地等于 ω_n[23]。因此，

$$\omega_{-3\text{dB}} = \frac{\sqrt{8}}{RC} \tag{129}$$

现在没有桥式 T-网络的带宽仅仅是 $1/RC$。因此，这个网络的带宽提升因子为$\sqrt{8}$，即大约 2.83，这已在这一章的主要部分提及。我们又一次看到，这个网络神奇地将带宽提升了将近 3 倍而没有增加功耗或者要求性能更好的晶体管。

12.8.5 最大平坦时延的设计公式

对于这个例子，我们再次需要零点与一个极点对相抵消以留下另一对复极点。唯一的差别是剩余的极点的衰减因子选择为对应最大时延平坦响应而不是最大幅度平坦响应。第一步，我们找出延时作为(负)相位函数的导数：

$$T_D = \frac{d}{dx}\left[\arctan\left(\frac{2\zeta x}{1-x^2}\right)\right] \tag{130}$$

这里 x 是归一化到 ω_n 的频率：

$$x \stackrel{\text{def}}{=} \omega/\omega_n \tag{131}$$

下一步，回忆起

$$\frac{d}{dx}(\arctan x) = \frac{1}{x^2+1} \tag{132}$$

那么

$$T_D = \frac{d}{dx}\left[\arctan\left(\frac{2\zeta x}{1-x^2}\right)\right] = \frac{2\zeta(-x^2+1)+4\zeta x^2}{4\zeta^2 x^2+(-x^2+1)^2} \tag{133}$$

计算一阶导数并稍作简化便得到

$$\frac{[4\zeta^2 x^2+(-x^2+1)^2][4(-\zeta x)+8(\zeta x)]-[2\zeta(-x^2+1)+4\zeta x^2][4(2x\zeta^2)-4(-x^2+1)x]}{[4\zeta^2 x^2+(-x^2+1)^2]^2} \tag{134}$$

这里观察到，在 DC 时它已经是零。

再看二阶导数，并令其直流下为零最终会得到

$$4\zeta-(2\zeta)(8\zeta^2-4)=0 \tag{135}$$

---

[23] 这个极点的星座图对应于二阶巴特沃兹低通滤波器的极点星座图。

解出得到衰减因子为

$$\zeta = \sqrt{3}/2 \tag{136}$$

相应的归一化的带宽为

$$\frac{\omega_{-3\mathrm{dB}}}{\omega_{\mathrm{n}}} = \left(\frac{\sqrt{5}-1}{2}\right)^{1/2} \approx 0.7862 \tag{137}$$

而衰减因子表示了一个电容比例，

$$\frac{C_{\mathrm{B}}}{C} = \frac{1}{12} \tag{138}$$

这样，联立这个比例的另一个表达式

$$\frac{C_{\mathrm{B}}}{C} = \frac{1}{4}\left[\frac{1-|k|}{1+|k|}\right] \tag{139}$$

得到耦合因子的大小为 1/2。

继续进行下去，我们得到自然频率为

$$\omega_{\mathrm{n}} = \frac{1}{R\sqrt{C_{\mathrm{B}}C}} = \frac{1}{R\sqrt{(C/12)C}} = \frac{\sqrt{12}}{RC} \tag{140}$$

计算带宽提升因子为

$$\frac{\omega_{-3\mathrm{dB}}}{1/(RC)} = \frac{\omega_{-3\mathrm{dB}}}{\omega_{\mathrm{n}}}\sqrt{12} \approx 2.72 \tag{141}$$

我们看到相对于最大增益平坦响应，我们仅付出了一个小的带宽损失的代价就获得了最大线性相移，总的来说，这个网络对元件值不是太敏感，所以不必对元件值的容差有太过严格的要求。

12.8.6 最大带宽的设计公式

看来推导过程将要完成了，现在给出最大带宽的设计公式，这仍然局限于完美的零极点抵消。

在这个例子里，我们首先发现对应于最大带宽的一个二阶系数表达式中的衰减系数。具体的推导过程没有太多指导意义，所以就简单地给出结果

$$\zeta = \left(\frac{5-\sqrt{15}}{2}\right)^{1/2} \approx 0.7507 \tag{142}$$

因此，

$$C_{\mathrm{B}} \approx C/9 \tag{143}$$

$$L_{\mathrm{pri}} \approx R^2C/1.229 \tag{144}$$

$$L_2 \approx 0.38538L_{\mathrm{pri}} \approx 0.31357(R^2C) \tag{145}$$

以 1dB 通带纹波为代价，带宽提升因子改善了 10%左右，差不多为 3。

从图 12.53 可以明显看出，这个选择是以增益和延时平坦度为代价而换取最大化带宽。由于这些有争议的增益和时延平坦度的变差仅仅是换取多了一点点的带宽，这种选择从没有在发表了的文献中实现过，甚至没有参考文献提到过它。我们这里提供这个数据只是为了使这个讨论更完整。

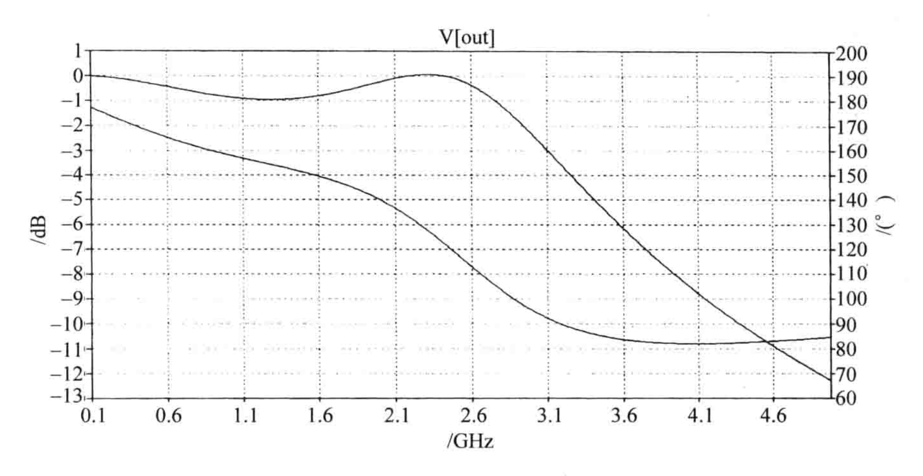


图 12.53　桥接 T 形线圈的 LTSpice 模拟结果（最大带宽情形）

12.8.7　总结

对于这个网络，作为两个最后的观察中的一个，有必要认识到对易的用途。尽管所有的推导假设输入/输出端口和变量已定，也可以反转这个假设而不改变传输函数。这样，可以通过用电流源驱动输出电容，而撮取出先前作为输入端口上的输出电压。如果晶体管的输出电容大于网络的负载电容，则这种对易是有用的。这样把晶体管的电容吸收到网络中就会有最大的效果。

最终，还可以从另一个角度来看这个电路，这样做是值得的。桥式 T 形网络可以被认为是集总传输线的模型的一个单元。在这种近似成立的场合下，延时单元的输入阻抗将在整个频带内呈现电阻性。这样，限制放大器带宽感的电抗性负载的影响将在很大程度上被抑制。

第13章 低噪声放大器(LNA)设计

13.1 引言

接收器的第一级一般是一个低噪声放大器(LNA),它的主要功能是提供足够的增益来克服后续各级(典型的是一个混频器)的噪声。除了提供这一增益而又附加尽可能小的噪声而外,LNA应当能处理大的信号而又不失真,并且常常还必须对输入信号源表现出一个特定的阻抗,例如50Ω。最后这一个考虑在LNA的前级是一个滤波器时特别重要,因为许多滤波器(无源或有源)的传输特性对于终断的质量是十分敏感的。

我们将会看到通过用一个其值取决于器件特性的神奇的信号源阻抗,可以从这个器件得到最小的噪声系数(noise figure,NF)。遗憾的是,这个源阻抗通常与使功率增益最大的阻抗值是不同的,也可能是完全不同的。因而伴随一个优异的噪声系数性能的也可能是差的增益与一个糟糕的输入匹配。这一章的一个目标就是在定量的基础上对此进行折中,以确保没有太痛苦的迭代就能得到满意的设计结果。

我们将主要集中在一个单端窄带LNA的电路结构,这个结构能够提供接近于最小的噪声系数,同时具有极佳的阻抗匹配与不错的功率增益。这种放大器窄带特性不一定就是一个包袱,因为许多应用反正要求有滤波过程。我们将要研究的LNA就具有这样一种对许多所希望是有的特性的一个平衡。

然而,在展开讨论之前,我们需要暂时偏离主题,而简略地用一般的术语来研究一下噪声问题。

13.2 经典的两端口网络噪声理论

在这一节中,我们给出一个两端口网络的噪声的宏观描述。针对这样的系统的噪声模型讨论能大大地简化分析,并且导致有用的设计内在见解的获取。

13.2.1 噪声因子

衡量对一个系统噪声性能的一个有用指标是噪声因子,通常表示为F。为了对它进行定义并了解为什么它有用,考虑一个含噪声(但是线性)的二端口网络,这个网络由一个输出导纳为Y_S及一个等效的并联噪声电流$\overline{i_s}$构成的噪声源驱动(见图13.1)。

$\overline{i_s}$　Y_S　带噪声的二端口网络

图13.1　为一个噪声源驱动的本身包含噪声的二端口网络

如果关心的只是整个的输入/输出特性,那么就没有必要非常复杂地去跟踪所有内部的噪声源。幸运的是,所有这些噪声源的总效应可以只用一对外部噪声源代

表：一个噪声电压和一个噪声电流。这一极大程度的简化使我们可以快速估计噪声源的导纳是如何影响总的噪声特性的。据此，我们能够确认一些准则，这些准则是为达到最优噪声性能所必须满足的。

噪声因子定义为

$$F \stackrel{\text{def}}{=} \frac{\text{总的输出噪声功率}}{\text{输入噪声源引起的输出噪声功率}} \tag{1}$$

式中按照惯例，噪声源处在 290K 的温度下①。噪声因子是对系统引起的信号噪声比变坏的一个度量。这一变坏程度越大，则噪声因子也就越大。如果一个系统没有加进任何它自己的噪声，那么总的输出噪声就完全取决于噪声源，于是噪声因子就为 1。

在图 13.2 的模型中，所有的噪声都表现为对无噪声网络的输入，所以可以在这个网络的输入端处计算噪声系数。直接基于公式(1)进行计算，要求我们计算出所有噪声源引起的总功率，然后再把这个计算结果除以由输入噪声源引起的功率。另一个等效的(更为简单的)方法是计算总的在短路条件下的均方噪声电流，然后把这个量除以由输入噪声源引起的短路均方噪声电流。这个替代的方法之所以等效是因为单个的功率贡献都正比于短路的均方电流，而其比例常数(它涉及到噪声源和二端口网络间的电流分流比)对于所有的各项都是相同的。

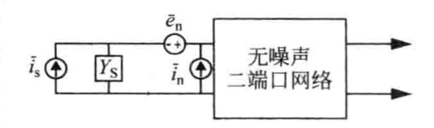


图 13.2　等效噪声模型

在进行这一计算时，我们通常遇到的问题是要把相互间具有不同程度相关性的噪声源合并起来。在相关性为零的特殊情况下，各个功率就可以简单叠加。例如，如果假设(这个假设看上去是合理的)噪声源和二端口网络的噪声功率完全不相关，那么噪声因子的表达式就成为

$$F = \frac{\overline{i_s^2} + \overline{|i_n + Y_s e_n|^2}}{\overline{i_s^2}} \tag{2}$$

注意，尽管我们已经假设了噪声源的噪声与二端口网络两个等效的噪声发生器是不相关的，但式(2)却并没有假设二端口网络的这两个噪声发生器也是相互间不相关的。

为了包括在 e_n 和 i_n 间可能相关的情形，把 i_n 表示成两个分量的和。一个分量 i_c 与 e_n 相关，另一个分量 i_u 则不相关：

$$i_n = i_c + i_u \tag{3}$$

由于 i_c 与 e_n 相关，因此可以把它看成是与 e_n 成比例，其比例常数的单位为导纳的单位：

$$i_c = Y_c e_n \tag{4}$$

这个常数 Y_c 称为关联导纳。

联立式(2)～(4)，噪声因子变成

$$F = \frac{\overline{i_s^2} + \overline{|i_u + (Y_c + Y_s)e_n|^2}}{\overline{i_s^2}} = 1 + \frac{\overline{i_u^2} + |Y_c + Y_s|^2\, \overline{e_n^2}}{\overline{i_s^2}} \tag{5}$$

① 你也许要问为什么把一个相对较冷的 290K 作为参照温度。其理由只不过是在那个温度下 $kT = 4.00 \times 10^{-21}$ J(焦耳)。就像许多操业的工程师们那样，Harald T. Friis 喜欢整数(见他的"Noise Figures of Radio Receivers"(无线电接收器的噪声系数)，*Proc. IRE*, July 1944, pp. 419-422)。他用 290K 作为参考温度的建议在那个用计算尺来进行计算的年代特别有吸引力。这个建议迅速被工程师们采用，并最终被各种标准化委员会所接受。

式(5)的表达式包含了三个独立的噪声源，每个都可以看成是由一个等效电阻或电导产生的热噪声（不管这样的电阻或电导实际上是否是噪声的来源）：

$$R_n \overset{\text{def}}{=} \frac{\overline{e_n^2}}{4kT\Delta f} \tag{6}$$

$$G_u \overset{\text{def}}{=} \frac{\overline{i_u^2}}{4kT\Delta f} \tag{7}$$

$$G_S \overset{\text{def}}{=} \frac{\overline{i_s^2}}{4kT\Delta f} \tag{8}$$

利用这些恒等式，噪声因子的表达式可以只用阻抗和导纳项来表示：

$$\begin{aligned} F &= 1 + \frac{G_u + |Y_c + Y_s|^2 R_n}{G_s} \\ &= 1 + \frac{G_u + [(G_c + G_s)^2 + (B_c + B_s)^2]R_n}{G_s} \end{aligned} \tag{9}$$

式中，已经明确地把每个导纳分解成电导 G 和电纳 B 的和。

13.2.2　最优的源导纳

一旦一个给定的二端口网络的噪声特性已用它的四个噪声参数（G_c，B_c，R_n 和 G_u）表示，式(9)就可以被用来求出使噪声因子达到最小的通用条件。求出对噪声源导纳的一阶导数并使其为零得到：

$$B_s = -B_c = B_{opt} \tag{10}$$

$$G_s = \sqrt{\frac{G_u}{R_n} + G_c^2} = G_{opt} \tag{11}$$

因此，为了使噪声因子最小，应当使噪声源的电纳等于关联电纳的负值，而噪声源的电导等于式(11)中的值。

相应于这一选择的噪声因子可以通过直接把式(10)和式(11)代入到式(9)中求出：

$$F_{min} = 1 + 2R_n[G_{opt} + G_c] = 1 + 2R_n[\sqrt{G_u/R_n + G_c^2} + G_c] \tag{12}$$

我们也可以用 F_{min} 和噪声源的导纳来表示噪声因子：

$$F = F_{min} + \frac{R_n}{G_s}[(G_s - G_{opt})^2 + (B_s - B_{opt})^2] \tag{13}$$

因此，噪声因子的等值线在导纳平面中是一些非重叠圆[②]，见图 13.3。

R_n/G_S 这个比值作为方程(13)的第二个项前面的一个乘子。对于一个固定的源电导，R_n 告诉我们噪声系数对偏离最优化条件的相对灵敏度。大的 R_n 意味着高的灵敏度；具有高的 R_n 的电路或器件迫使我们更卖力气地工作以确定、取得及维持最优化条件。我们很快就会看到工作在低偏置电流下的工作状态导致大的 R_n，这与我们有关当功耗预算是紧的时候，取得高性能是更为困难的一般直觉是一致的。

---

② 当画在史密斯图上时它们也是圆，因为在这两个坐标平面间的映射是双线性变换，因而保持了圆的形状。

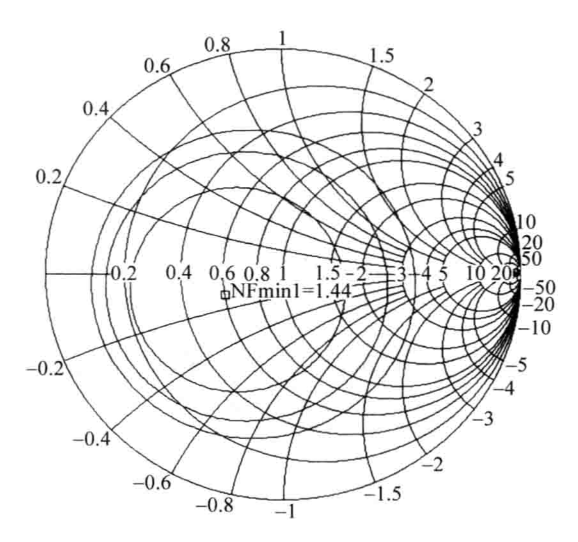


图 13.3　噪声系数圆的例子(根据 AppCAD 软件对 Agilent ATF-551M4 pHEMT 的模拟结果)

重要的是要认识到，尽管使噪声因子最小化在一定程度上有助于最大功率的传输，但使这两个条件成立的源导纳一般并不相同，这可以从式(10)和式(11)中明显看出。例如，没有任何理由期待关联电纳一定等于输入电纳(除非是出自偶然)。因此一般地说，如果要使噪声性能最优，那么必须接受小于最大功率增益这一事实，反之亦然。

13.2.3　经典噪声优化方法的局限

上面介绍的经典理论隐含地假设了给我们的器件具有特定的固定特性，然后定义源导纳以使得噪声系数最小。这在分立器件 RF 设计中是常见的。然而，经典的优化方法并不能直接容纳集成电路(IC)实现中器件尺寸能被选择的这样一个自由度。这样，为了进行合理的某种优化，在经典优化方法之外的考虑需要被提供。因为我们这里的注意力是在分立元件电路，我们不进一步讨论这个命题。之所以提到这个情况是为了指出经典方法的不完整性③。

13.2.4　噪声系数与噪声温度

除了噪声因子外，在文献中常出现的其他的品质因子为噪声系数和噪声温度。噪声系数只不过是用分贝表示的噪声因子④。

噪声温度 T_N 是一个表示放大器噪声贡献的另一种方法，它定义为引起在参照温度 T_{ref}(即 290K)下的全部输出噪声所需要的源电阻的温度上升。它与噪声因子的关系如下：

$$F = 1 + \frac{T_N}{T_{ref}} \Rightarrow T_N = T_{ref}(F - 1) \tag{14}$$

一个没有附加它本身噪声的放大器具有的噪声温度为绝对零度(0K)。

③　关于在 CMOS IC 环境下这个课题的更多讨论，见 T. Lee. *The Design of CMOS Radio-Frequency Integrated Circuits*(CMOS 射频集成电路的设计)，2nd ed.，Cambridge University Press，2004。

④　使事情复杂化的是，在某些教科书中，噪声因子和噪声系数的定义恰好是互换的。

正如我们在这一章后半部可以看到的那样，噪声温度在描述级联放大器性能时特别有用，或者是对那些噪声因子非常接近 1(或噪声系数非常接近 0dB)的放大器也特别有用。因为在这些情形中，噪声温度对噪声性能提供了一个更高分辨率的描述。这可以从表 13.1 中看出。范围在 2～3dB 的噪声系数一般被认为很好，而在 1dB 左右或以下的值则被认为非常好。

表 13.1　噪声系数、噪声因子和噪声温度

NF(dB)	F	T_N/K
0.5	1.122	35.4
0.6	1.148	43.0
0.7	1.175	50.7
0.8	1.202	58.7
0.9	1.230	66.8
1.0	1.259	75.1
1.1	1.288	83.6
1.2	1.318	92.3
1.5	1.413	120
2.0	1.585	170
2.5	1.778	226
3.0	1.995	289
3.5	2.239	359

13.3　双极型噪声模型的推导

在我们能客观地评估窄带低噪声放大器(LNA)电路拓扑的优点(以及局限性)之前，有必要首先对双极型晶体管来推导一个恰当的噪声模型。为了使得分析一步步进行，而且为了得到设计的直觉，我们将需要做一些简化的假设。这些假设在器件工作在截止频率 f_T 之下很多(如至少是 1/5 或更小)时不会引起太多的误差。在比这个频率更高的频率下，其他器件的特性(如增益)快速变差，这使得在这些频率下的噪声分析变得不必要，无论是准还是不准。

在一个双极型晶体管中的两个结，每一个都产生散弹噪声(shot noise)，这个噪声用一个并联的电流源来建模，噪声电流源的均方谱密度是 $2qI_{DC}$，其中 I_{DC}是通过该结的偏置电流值(参见图 13.4，译者注)。来自两个结的散弹噪声电流对大多数实际的应用都可以处理成不相关联，因此在下面的讨论中，忽略它们间的关联性。这个忽略可以允许我们将噪声功率直接相加起来。即，通过引用噪声源的统计独立性，一种有趣(且十分有用)的叠加原理成为了可能。

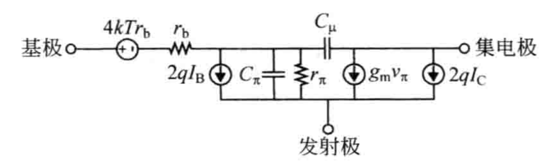


图 13.4　双极型晶体管的噪声模型

除了散弹噪声成分(这在某种意义上是带有根本性质的，因为无论在器件设计中是如何动脑筋，都无法消除这种噪声)之外，也存在着热噪声源：串联的基极电阻 r_b。这个噪声用有关串联的电压源来代表，其均方密度为 $4kTr_b$。在现代的器件中，这个噪声相比于发射极或集电极电阻的噪声通常是占主要地位(而且要高出不少)，因此可以忽略因发射极与集电极电阻带来的噪声。就如我们将要看到的，r_b 是非常不希望的。除了产生噪声(因而使噪声系数变坏)外，它的存在往往将产生最小噪声系数的源电阻抬高到不方便实现的值。

尽管会产生这样的想法，即出现在晶体管模型中的所有电阻器（如 r_b）都会贡献热噪声，但这样做等于双重计算。比如，r_π 源于线性化结的行为，而这个结的噪声已被散弹噪声建模了。因此，在产生于这种线性化的电阻与那些本来就是普通电阻器的电阻之间存在着差别。前者并不产生热噪声，而后者则是。

最后，集电极-发射极输出电阻通常（但并不是总是）大到这样的程度，使得它们在高的工作频率下可以被忽略，所以我们在所有以后的分析中都加以忽略[⑤]。

基于上面这些考虑的一个小信号晶体管模型在图 13.4 中给出。这个模型，尽管看起来简单，是抓住了为计算一个双极型晶体管放大器噪声系数的最重要的效应。这个模型对推导一个放大器的噪声系数有用的准确表达式是足够了，对于确定最佳的源电阻的目的也够用了。

译者注：重复定义式(1)为式(15)

$$F \stackrel{\text{def}}{=} \frac{\text{总的输出噪声功率}}{\text{输入噪声源引起的输出噪声功率}} \tag{15}$$

为了用表达式(15)，即(1)，计算噪声因子，将一个（带有热噪声）的源电阻连到图 13.4 的电路，如何进行计算，见图 13.5。注意，这个电路的输出端口是交流短路的。当然，在一个实际的电路中，输出端口会用其值不为零的电阻器作为负载——除非你的目标是在创造一个高科技的空间加热器（作者的玩笑话，译者注）。但是从式(15)可以清楚看出一个集电极负载电阻作为一个乘因子既出现在分子也出现在分母中。因此，结果是这个负载电阻最终被抵消了，所以无论什么电阻值对我们来讲都一样。因此，我们选择了负载电阻值为零，这是一个特别方便的值。

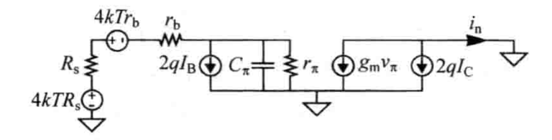


图 13.5　计算噪声系数的模型

一个很大胆的技巧使我们已经任意地消除了集电极-基极电容。这个电容的存在使得整个分析是如此之复杂，以至于将其移去对于即使是仅为了清楚起见都是十分有必要的。只要集电极负载是一个低阻抗，整个忽略通常不会是那么严重。然而，在一般的情况下，在那里需要考虑任意大小的集电极负载，忽略 C_μ 会产生严重的误差。最大的误差发生在计算对应于最小噪声系数的源电阻时。取决于负载阻抗的具体特性，这个最佳的源电阻可以变得过大或过小。幸运的是，那个最小的噪声系数倒是没有被改变太多，因此，我们将依然这样来推导噪声系数，前提是我们充分知道这样做后面所做的假设。

给定这些假设并采用短路负载，这个电路的噪声因子就是流经集电极支路的两个短路电流（标为 i_n）的比。其分子是因为所有的噪声源引起的均方短路电流的和，而分母是只因源噪声的均方短路电流。因此，有

$$F = \frac{2qI_C + 2qI_B\,|\,z_\pi \,||\, (r_b + R_S)\,|^2 g_m^2 + (4kTr_b + 4kTR_S)\left|\dfrac{z_\pi}{R_S + r_b + z_\pi}\right|^2 g_m^2}{4kTR_S\left|\dfrac{z_\pi}{R_S + r_b + z_\pi}\right|^2 g_m^2} \tag{16}$$

---

⑤　集电极-发射极电阻对厄利；它不产生热噪声，因为它是对结的宽度变动效应线性化的结果。最后，存在着集电极与基极间的反馈电阻，这是源于基区宽度调制。这个电阻在 RF 工作下，几乎总是可以完全被忽略。

其中,在分子上的简单求和是忽略在噪声产生器之间的任何相关性的直接结果。

让我们逐项来研究这个方程以理解其代表的意思,这看来是一个好的主意。在分母中,源电阻噪声均方电压谱密度 $4kTR_S$ 首先被乘以电压分压因子幅度的平方,以得到 r_π 两端的均方电压。这个平方了的电压又被乘以跨导 g_m 的平方,来得到平方的集电极电流,这样就完成了得到分母的工作。

从左到右考察分子中的各项,注意电阻器 r_b 的噪声电压发生器与电阻器 R_S 的噪声电压发生器是串联的。因此,两个噪声电压发生器同样是串联变换的,这解释了为什么分子中三个相加项的最后一项是现在这个样子的。

那个基极散弹噪声电流看到的是一个总的阻抗,这是 z_π(其本身又是 r_π 与 C_π 的并联)并联上(r_b+R_S)。用这个阻抗的平方幅度去乘均方散弹噪声电流就给出了 r_π 两端的均方电压。再将这个量乘以 g_m 的平方就得到了基极散弹噪声对均方集电极电流的贡献。

最后,集电极散弹噪声不需经任何乘子或其他的变换,这个项就直接加到在分子中的其他贡献项。

方程(16)可以通过消除几个共同项可以简化来得到(需要重新排一下序)

$$F = 1 + \frac{r_b}{R_S} + \frac{2qI_C + 2qI_B \mid z_\pi \mid\mid (r_b + R_S) \mid^2 g_m^2}{4kTR_S \mid z_\pi/(R_S + r_b + z_\pi) \mid^2 g_m^2} \tag{17}$$

进一步简化为

$$F = 1 + \frac{r_b}{R_S} + \frac{2qI_C \mid R_S + r_b + z_\pi \mid^2}{4kTR_S \mid z_\pi \mid^2 g_m^2} + \frac{2qI_B(r_b + R_S)^2}{4kTR_S} \tag{18}$$

还能继续消去共同项可得到更为简单的形式:

$$F = 1 + \frac{r_b}{R_S} + \frac{\mid R_S + r_b + z_\pi \mid^2}{2R_S \mid z_\pi \mid^2 g_m} + \frac{(r_b + R_S)^2 g_m}{2\beta_F R_S} \tag{19}$$

在得到这最后一个表达式时,我们已经用到双极型晶体管的跨导是 $qI_C/(kT)$及集电极与基极电流之比是 β 这些事实。

注意,第二项计入了由基极电阻直接引起的噪声,第三项是因为集电极散弹噪声,然后最后那一项是基极电流散弹噪声。这就是整个公式的最后形式,它使得我们可以对各项进行来源的确定。

同时注意到,表达式(19)包含有三类项(在所有的括号项都被乘出来后):一个是独立于 R_S 的项,另一类是正比于 R_S,第三类是反比于 R_S。对于非常小的源电阻,反比项占主要地位,而在源电阻值非常大时,正比项起主要作用。这样,在源电阻“非常小”与“非常大”之间,存在着一个使得这个和(亦即噪声因子)最小的优化值。在计算这个优化值之前,让我们从直觉上理解一下,为什么一个优化的 R_S 值会存在。

在非常低的源电阻时,基极电阻的贡献相比于源电阻的贡献更大,因此噪声系数变坏。

在非常高的源电阻时,由于基极散弹噪声造成的输出噪声贡献比较大(因为散弹噪声源面对的阻抗更大了,这样就产生了 r_π 两端更大的电压,进而从集电极出来更大的电流)。同时,由于源本身的输出噪声变小了,这是因为被 R_S 看到的电压分压器不利于源噪声进入晶体管的基极(然后被放大)。至于集电极散弹噪声的幅度则没有变化,但是它相对于 R_S 的贡献变坏了(即变大,译者注),这样噪声系数进一步变差。这样,R_S 的最优值在基极电阻的贡献与基极及集电极散弹噪声的效应间起到平衡作用。

我们将用到的噪声因子表达式是式(19)的稍微扩展的版本:

$$F = 1 + \frac{r_b}{R_S} + \frac{(R_S + r_b)^2 + |z_\pi|^2 + 2(R_S + r_b)\mathrm{Re}\{z_\pi\}}{2R_S |z_\pi|^2 g_m} + \frac{(r_b + R_S)^2 g_m}{2\beta_F R_S} \tag{20}$$

现在让我们进行数学运算以推导 R_S 的最佳值。

最佳的源电阻

发现这个最佳值的步骤是十分直截了当的。对源电阻取一阶导数，令其为零，然后希望得到解就是对应着最小值(而非最大值)：

$$\frac{\mathrm{d}}{\mathrm{d}R_S}\left(1 + \frac{r_b}{R_S} + \frac{|R_S + r_b + z_\pi|^2}{2R_S |z_\pi|^2 g_m} + \frac{(r_b + R_S)^2 g_m}{2\beta_F R_S}\right) = 0 \tag{21}$$

一直向着得到解而努力计算，得到了如下的一系列运算式：

$$\frac{\mathrm{d}}{\mathrm{d}R_S}\left(\frac{r_b}{R_S} + \frac{(R_S + r_b)^2 + |z_\pi|^2 + 2(R_S + r_b)\mathrm{Re}\{z_\pi\}}{2R_S |z_\pi|^2 g_m} + \frac{(r_b^2 + R_S^2 + 2r_b R_S) g_m}{2\beta_F R_S}\right) = 0 \tag{22}$$

$$\frac{\mathrm{d}}{\mathrm{d}R_S}\left(\frac{r_b}{R_S} + \frac{R_S^2 + r_b^2 + |z_\pi|^2 + 2r_b \mathrm{Re}\{z_\pi\}}{2R_S |z_\pi|^2 g_m} + \frac{(r_b^2 + R_S^2) g_m}{2\beta_F R_S}\right) = 0 \tag{23}$$

在式(23)中，我们已经抛除了那些独立于 R_S 的项，因为它们对 R_S 的微分为零(我们在得到式(19)时已经弃除了一个单位相加因子，假若你曾经在想这一项到哪里去了)。如果我们只想去用最后的答案而不是经历这个推导的每一个步骤，可以跳过这个过程。

将正比于 R_S 的项与反比于它的项分开，得到

$$\frac{\mathrm{d}}{\mathrm{d}R_S}\left[\frac{1}{R_S}\left(r_b + \frac{r_b^2 + |z_\pi|^2 + 2r_b \mathrm{Re}\{z_\pi\}}{2|z_\pi|^2 g_m} + \frac{r_b^2 g_m}{2\beta_F}\right) + R_S\left(\frac{1}{2|z_\pi|^2 g_m} + \frac{g_m}{2\beta_F}\right)\right] = 0 \tag{24}$$

现在可以取微分，并将其设为零，得到

$$\left(\frac{1}{R_S^2}\right)\left(r_b + \frac{r_b^2 + |z_\pi|^2 + 2r_b \mathrm{Re}\{z_\pi\}}{2|z_\pi|^2 g_m} + \frac{r_b^2 g_m}{2\beta_F}\right) = \frac{1}{2|z_\pi|^2 g_m} + \frac{g_m}{2\beta_F} \tag{25}$$

因此，这个最佳的源电阻(平方后)为

$$R_S^2 = \frac{r_b + \dfrac{r_b^2 + |z_\pi|^2 + 2r_b \mathrm{Re}\{z_\pi\}}{2|z_\pi|^2 g_m} + \dfrac{r_b^2 g_m}{2\beta_F}}{\dfrac{1}{2|z_\pi|^2 g_m} + \dfrac{g_m}{2\beta_F}} \tag{26}$$

略微加以简化，得

$$R_S^2 = \frac{2|z_\pi|^2 g_m r_b + r_b^2 + |z_\pi|^2 + 2r_b \mathrm{Re}\{z_\pi\} + g_m^2 r_b^2 |z_\pi|^2/\beta_F}{1 + |z_\pi|^2 g_m^2/\beta_F} \tag{27}$$

式(27)是不加任何附加的近似可以直接追溯到我们的噪声模型的最后形式。然而，如果允许一个或两个非常合理的近似，进一步的简化是可能的。一个假设是工作频率远高于 $1/r_\pi C_\pi(=\omega_T/\beta)$，但是又没有高到使集总参数模型失效。另一个假设是偏置电流足够高以至于 C_π 主要由扩散电容决定。有了这些假设，可以写出

$$R_S^2 \approx \frac{\left(\dfrac{\omega_T}{\omega}\right)^2\left(\dfrac{2r_b}{g_m} + \dfrac{r_b^2}{\beta_F} + \dfrac{1}{g_m^2}\right) + r_b^2 + \dfrac{2r_b r_\pi}{(\omega/\omega_T)^2 \beta_F^2}}{1 + \dfrac{(\omega_T/\omega)^2}{\beta_F}} \tag{28}$$

假若这是通常发生的，分子上的最后一项比起它前面那一项要来得小，那么

$$R_S^2 \approx \frac{\left(\dfrac{\omega_T}{\omega}\right)^2\left(\dfrac{2r_b}{g_m} + \dfrac{r_b^2}{\beta_F} + \dfrac{1}{g_m^2}\right) + r_b^2}{1 + \dfrac{(\omega_T/\omega)^2}{\beta_F}} \tag{29}$$

作为一个具体的数字例子，考虑在 1GHz 下用一个型号为 2SC3302 的微波晶体管。假

设集电极电流是 10mA，在这个偏置下，跨导是 400mS，$\beta=80$，并且 $\omega_T=10\pi\text{Gr/s}$。剩下的一个未知量是 r_b 的值，它可能还是不知道因为数据手册很少给出其值(2SC3302 依然不是例外)。然而，可庆的是，一个输入阻抗与频率关系的图是给出的。这个图表明，在偏置电流为 20mA 时(在此条件下，利用其他数据手册给出的信息可以估计 C_π 大约为 23pF)，在大约 800MHz 处存在着谐振。这个谐振是封装和引线的电感与 C_π 互相作用的结果。在那个谐振条件下，输入电阻大约在史密斯圆图 25Ω 与 50Ω 两个等电阻圆中间左右的位置，因此可以估计总的电阻约为 37～38Ω。这个电阻是 r_b 与由串联发射极电感产生的阻抗的实部项的和。而这个串联电感是与封装及引线相联系的。就如在下一节中会证明的，这个诱发的电阻其值为 $\omega_T L_e$。这个寄生电感并不容易被估计出来，但是我们能够从谐振频率计算出总的电感大约为 1.7nH。这个值从封装的物理尺寸来看是相当可信的。假设这个总的电感在基极与发射极之间对分(尽管实际上并不是如此)，这允许我们来估计由这个诱生电阻对总电阻的贡献大约为 30Ω。因为这个值是如此接近总的被估计的输入电阻，我们关于 r_b 值的不确定性是大的。然后，我们还是这样做了，认为 r_b 的值是 7～8Ω。

在这些条件下，最佳的源电阻大约是 35Ω(此时，1GHz 下的噪声系数是 2dB)。这个值十分接近 50Ω，因此即使不采用阻抗变换，也只会引起不大的 NF 的惩罚(比 0.1dB 大一点)。当偏置电流为 5 mA 时，噪声系数会增加，同时对用 50Ω 的源电阻带来的惩罚也增加了(大约 0.2dB)。总的用 50Ω 源电阻的最小噪声系数是大约 3dB(再一次说明，这个值是在工作频率为 1GHz 的)。

作为对我们的推导的检查，比较一下计算得到的 2dB 噪声系数与数据手册在 1GHz 下给出的 1.7dB 的最小值。对 500MHz 的工作频率重复我们的计算，得到了 1.6dB NF(在 5mA 偏置)，这个值可以与同样条件下数据手册给出的 1.5dB 的值相比较。考虑到我们所做的近似于参数提取的粗略的性质，总的符合程度是令人满意的。

最后，记住一个放大器必须要能放大。取得一个低的噪声系数是重要的，但这只是努力的一半。因为这个原因，选择一个适当的偏置必须同时考虑增益与噪声系数。在 2SC3302 这个特定的情形下，较大的偏置电流会带来较高的增益，这主要是因为在大的偏置电流下，f_T 接近于最大值。因为最小的可取得的噪声系数并不在我们考虑的偏置电流范围内变化剧烈，存在着相当大的设计空间，因此其他的因素(比如电路的实现)也可以一并加以考虑。

13.4　窄带 LNA

前面一节的推导表明产生最小噪声因子的源阻抗一般与使功率传输最大的条件不相关联。而且，双极型晶体管的高频输入阻抗从本质上讲是电容性的，所以提供一个对 50Ω 信号源的良好匹配而又不降低噪声性能看起来是比较困难的。由于对外部电路表现为一个已知的电阻性阻抗几乎总是对一个 LNA 的关键性要求，所以在设计中也要强加这个要求。

产生一个实数的输入阻抗而又不让噪声特性变差的一个特别好的方法是采用电感性的发射极简并。这个方法起源于 20 世纪 30 年代的模拟信号真空管放大器。它对基于场效应管(FET)的放大器一样有效。因此，除了为满足偏置电路差异而作的微小的修正，下面的描述可以被理解为也适用于用其他类型的晶体管构建的 LNA。

当发射极有了一个简并电感，基极电路除了经受一个因输入电容而引起的正交相位偏

移外，还有额外的相移，从而在输入阻抗里出现了一个电阻分量。这个方法的一个重要优点是可以通过选择电感来控制阻抗实部的值，这从计算图 13.6 所示的电路的输入电阻中清楚看出。

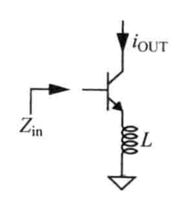


图 13.6 电感性简并了的共发放大器

为了简化分析，考虑一个仅包括跨导与基极-发射极电容的器件模型。在这个情形下，不难证明输入阻抗有如下形式：

$$Z_{in} = sL + \frac{1}{sC_\pi} + \frac{g_m}{C_\pi}L \approx sL + \frac{1}{sC_\pi} + \omega_T L \tag{30}$$

因此，输入阻抗是一个串联 RLC 网络的阻抗，其电阻项直接正比于电感的值。

更一般化地说，一个任意的发射极简并阻抗 Z 当发射到基极电路上时，被一个等于 $[\beta(\mathrm{j}\omega)+1]$的因子修正，这里 $\beta(\mathrm{j}\omega)$是电流增益：

$$\beta(\mathrm{j}\omega) = \frac{\omega_T}{\mathrm{j}\omega} \tag{31}$$

这个电流增益在 ω_T 时，其幅度变成单位值 1，这是意料中的，并且因为 C_π 有一个电容的相角。因此，在一般情况下，

$$Z_{in}(\mathrm{j}\omega) = \frac{1}{\mathrm{j}\omega C_\pi} + [\beta(\mathrm{j}\omega)+1]Z = \frac{1}{\mathrm{j}\omega C_\pi} + Z + \left[\frac{\omega_T}{\mathrm{j}\omega}\right]Z \tag{32}$$

注意，一个电容性的简并会对输入阻抗关系一个负的电阻[⑥]。因此，如何从发射极到地的寄生电容会抵消由电感性简并产生的正电阻。在任何实际的设计中，考虑到这个效应是十分重要的(或者是设法去利用这个效应，为你的目的服务)。

不管这个电阻项的值是什么，重要的是要强调它并不带来一个普通的电阻器具有的热噪声，因为一个纯的电抗是不会产生噪声的。我们因此可以利用这个性质来提供一个特定的输入阻抗，而不会使放大器的噪声性能变坏。

然而，式(30)清楚地表明这个输入阻抗只在一个频率点上(即谐振时)是纯电阻性的。幸运的是，在许多场合，当窄带操作不仅是可以被接受的而且是被希望的，这种电感性简并是一个十分有价值的技术。我们在这一章的余下部分将要研究的 LNA 拓扑因此如图 13.7 所示。

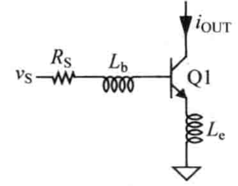


图 13.7 带有电感性发射极简并的窄带 LNA (偏置未画出)

电感 L_e 是被选择来提供所要求的输入电阻(等于源电阻 R_S)。因为输入阻抗只在谐振时为纯电阻性的，需要有一个附加的自由度，由电感 L_b 提供，来保证这个条件[⑦]。现在，在谐振时，基极对发射极的是输入电压的 Q 倍。在这个条件下，这一级电路总的跨导 G_m 因此为

$$G_m = g_{m1}Q_{in} = \frac{g_{m1}}{\omega_0 C_\pi(R_S + \omega_T L_e)} = \frac{\omega_T}{2\omega_0 R_S} \tag{33}$$

式中，我们已经用了 ω_T 是 g_{m1}对 C_π 的比这个近似。

⑥ 以往电容性负载的射极跟随器的不稳定性，名声不好。这个负的输入电阻是在根本上要负责的，并且解释了为什么在基极或栅电路上加上某些串联的正电阻可以帮助解决这个问题。

⑦ 也有可能封装及其他的寄生电感提供了比这个要求的电感值还要大的电感。在这个情况下，可能需要一个串联的电容来在要求的频率点上使得输入回路谐振。

设计的步骤因此是相当直截了当的。首先选择一个与增益及噪声系数目标一致的偏置电流。然后，计算最佳的源电阻来使噪声系数最小。接着是加上足够的发射极简并电感来产生一个输入阻抗，其实部等于最佳的原电阻，再在基极电路上加入足够的类型合适的阻抗（比如，更多的电感）来移去残余的电抗性输入成分，这样就将输入回路置于谐振。最后，在实际的信号源与放大器间插入一个不带损耗的匹配网络（如果需要的话）将 50Ω（或其他的源电阻值）转换到 R_S 的最佳值（这一段描述不甚确切，其实就是进一步阻抗变换的网络，译者注）。这个匹配网络经常能合并到需要使输入回路谐振的所要求的电感（比如说，因为也可能是电容，译者注）。

这个特定的步骤是十分吸引人的，因为它对所有涉及到的参数都进行了平衡。一个完美的匹配通过电感性简并得到保证，而且这个技术提供了在给定的偏置条件下最低可能的噪声系数。在输入电路上的谐振条件也同时确保了高的增益，这是因为有效的电路级的跨导正比于 ω_T/ω。

前面的分析对于突出这个电路的一阶行为是足够了。更详细的分析可以涵盖其他效应，例如有限大小的晶体管输出电阻 r_0。给出图 13.7 所示的电路的输入阻抗是相当简便的，

$$Z_{in}(j\omega) = \frac{1}{j\omega C_\pi} + j\omega L + g_m \frac{L}{C_\pi}\left(\frac{r_0}{r_0 + j\omega L + Z_L}\right) \tag{34}$$

其中 Z_L 是加到集电极上的阻抗。相比于我们前面方程的结果可以看到，有限大小的输出电阻改变了在阻抗公式中的第三项。特别是，我们看到圆括弧中的这一项仅在极限情况下为单位值。一般情况下，Z_{in} 的实部会被减小，而虚部也会发生改变，因而造成输入回路谐振频率的偏移。假若这个情形是十分普遍的，负载是一个并行谐振糟路，那么 $|Z_L + j\omega L|$ 这个量可能在谐振频率上或附近是足够的大（相对于 r_0 而言），从而引起输入阻抗实部的一个大的跌落。取决于输入与输出回路的相对谐振频率，足够实部的跌落有可能发生在低于、正好或高于整个放大器希望工作的中心频率。不言自明，这个跌落的大小与位置是最需要关心的。假若跌落的发生远离所希望的工作频率，它的存在可能不会是什么大问题。但是，通常跌落是发生在中心频率附近的几个百分点处（因为输入与输出电路的谐振频率是设计成互相接近的），结果就在感兴趣的频段内某处产生了差的输入匹配。

一个可能的解决方案是采用共射共基级。然而，这个方法只是部分有效，因为同样因为这个在最初造成问题的 r_0，也限制了共射共基电路结构的有效性。在十分难解决的场合，可能有必要去采用基几级堆叠的共基的共射共基级。降低集电极负载的负载电阻也经常会产生令人满意的设计。实际上，单单是知道上面这个问题本身，就常常能避免发生不佳的，让人惊讶的设计结果。

13.5　几个实用的设计细节

13.5.1　实现发射极简简并电感

窄带 LNA 依赖于电感性的发射极简并来产生输入阻抗中的一个实数项。在微波频率下，经常会发生所需要的值是如此之小以至于十分困难来制作。举一例，还是继续用我们的 2SC3302 晶体管，偏置在 10mA 下，我们会需要大约 2nH 来产生 50Ω。而如果偏置电流减小到 5mA，这个要求的电感值是上面值的 2～3 倍（为取得最小 NF，阻抗的目标值要增加，

加上 ω_T 值的减小，都会引起所需要的电感值增加得比没有考虑到这些因素时的期待值要快）。然而，2nH 是我们若不加特别小心的情况下能得到的最小值，同时要考虑到这个值的相当一部分已经被包括在封装内了。所以，要控制制作精确电感值是十分具挑战性的。在封装与引线电感已经超出所需要的值的情况下，输入阻抗会实际上表现出电感性的，此时要使输入回路谐振，需要一个电容（而不是电感）。为了避免产生这样的情况，在画版图与构筑时，要特别加以小心。

13.5.2 集电极负载

在通常的情形下，希望有一个谐振糟路作集电极负载。这样的一个负载通过吸收任何存在的输出电容来增加增益。而且，附加的对不想要的信号的滤波是十分希望的。

存在着几个实现这样的负载的实际可行的途径。其中之一是用分立元件的具有恰当值的电感器。对于我们的目标来讲，用一根适当长度的微带线来实现电感器，更为合适。这是因为微带线的多样性，见图 13.8。

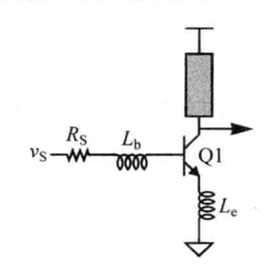


图 13.8　带有微带线负载的窄带 LNA（偏置未给出）

微带线的长度被调整到来产生谐振。而且，如果需要的话，一个阻抗下变换的实现很容易通过沿着这根线在一个中间的位置抽出头来得到输出。很明显，阻抗值在线的 V_{CC} 端最小，而在集电极端最大。作为一个一阶（并且是粗糙）的近似，阻抗沿着线呈平方率关系变化。

这个谐振峰的宽窄可以通过改变线的宽度来调整。这个宽度控制着线的 L/C 的比，因而也就控制了 Q。微带线的其他吸引人的特性是它使得实现偏置相对说来容易些。这会在下一节中看到。

13.5.3 偏置电路

在低频工作下，有许多方法可以来给单端放大器提供偏置。然而在微波频率下，我们可选的偏置方法就没有那么多了，这是因为我们不能总是允许使用那些在偏置网络的分立元件实现中采用的阻抗值。比如，在低频下通过一个分压器来给基极加偏置，然后插入一个起稳定作用的发射极简并电阻，这是十分经常的做法。为了补偿失去的增益，通常将一个旁路电容器跨接在这个简并电阻器两端。

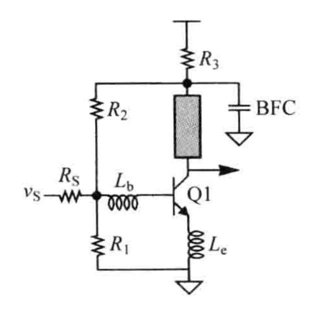


图 13.9　通过 V_{BE} 乘法器来进行偏置的窄带 LNA

在现在的情况下，采用这个方法大概是不实际的，因为在发射极电路中的任何不需要的元件只会使得我们在那里实现值很小的电感更加困难。然而，因为加发射极电阻性的简并元件的目的就是为了通过负反馈来降低直流（DC）增益，可以寻求完成同样净目标的替代方法。我们能够加一个从收集极到基极的负反馈，如图 13.9 所示。

这个电路的详细分析留给读者作为一个练习，但是一个快速的定性描述是 R_1 两端的电压 V_{BE}。如果可以忽略基极电流，那么流过 R_1 与 R_2 的电流是一样的。结果是，R_2 两端的电压是 R_1 两端电压的乘数倍，因而是 V_{BE} 的乘数倍。

而直流输出电压则为

$$V_{\mathrm{OUT}} = V_{\mathrm{BE}}(1 + R_2/R_1) \tag{35}$$

因为 V_{BE}是对温度敏感的，因而输出电压也是如此。然而对于我们的目的而言，这个变化是足够小的，因此这依然是一个有用的电路。

集电极负载 R_3 是由一个电容器旁路的，因此作为负载的微带线的顶端保持校好的信号接地。然而，这个旁路不需要是非常好的，因为不良接地造成的附加的电感这里仅迫使微带线要缩短一点。将偏置电路反馈网络的始发点由发射极转移到集电极就这样解决了这个头痛的问题。

作为有关偏置方法的最后一个注解，如果偏置点要保持基本上不受偏置电流变化的影响，那些电阻器的值要选得足够小以使得流经它们的电流与晶体管的基极电流变化相比是大的。这个要求与希望保持电阻器的值足够大以使得它们对热噪声的贡献最小化相矛盾。幸运的是，通常并不困难来找到一个平衡点可以使噪声系数的变坏被保持在几个十分之一分贝，或更少。

13.6　线性度与大信号性能

除了噪声系数、增益与输入匹配，线性度也是一个重要的考虑，这是因为 LNA 必须要在不增加过多的噪声的前提下完成比简单的放大更多的功能。当 LNA 接收强信号时，它必须保持足够的线性度。特别是，LNA 在一个强的干扰信号存在的情形下接收弱信号时，必须保持线性地工作，不然的话各种各样的电路工作的病态状况会出现。这些交调失真的后果包括失去敏感度（也被称为阻塞）与跨调制。阻塞发生在当因强的干扰信号引起的交调乘积项远远大于所希望接收的弱信号，而跨调制发生在放大器的非线性交互作用将一个信号上的调制信号转移到另一个信号的载波上去。当然，这两个效应都是不希望的，因此 LNA 设计者的另一个任务是将这些问题控制在最小的实际影响程度。

在本章所描述的 LNA 设计步骤并不直接讨论线性度，因此我们发展一些评估放大器大信号性能的方法，重点是获取设计的直觉内在理解。就如我们将要看到的，尽管窄带 LNA 电路拓扑在多少牺牲线性度的代价下取得好的噪声性能，这个折中并不那么严重，以至于阻止实现能满足性能要求高的应用的足够的动态范围。

尽管有许多度量线性度的方法，但最普遍使用的是三阶交调（IP3）和 1dB 压缩点（$P_{1\mathrm{dB}}$）[⑧]。三阶交调点作为一个线性度的量度是在 Avantek 大约于 1964 年首次提出的。为了使这两种度量与容易计算得到的电路和器件参数联系起来，假设放大器的输出信号可以用一个幂级数来表示[⑨]。进一步假设我们是用足够小的信号来估计这些线性度的度量的，因此当舍去级数中三次项以后的各项时引起的误差可以忽略不计：

$$i(V_{\mathrm{DC}} + v) \approx c_0 + c_1 v + c_2 v^2 + c_3 v^3 \tag{36}$$

这个表达式描述了跨导这一特定情形。

---

⑧　在直接变换（即自差）接收机中，二阶交调更为重要。

⑨　我们也假设输入和输出间是以无滞回线（无记忆）的过程联系在一起的。一个更为精确的方法是采用如 Volterra 级数，但其结果很复杂使我们无法得到我们正在寻求的对许多设计问题的内在理解。

现在考虑幅值相同频率稍有不同的两个正弦输入信号：

$$v = A[\cos(\omega_1 t) + \cos(\omega_2 t)] \tag{37}$$

把式(37)代入到式(36)中，经过简化和项的合并，我们就可以确定输出频谱中的分量[⑩]。DC 和基波分量如下：

$$[c_0 + c_2 A^2] + \left[c_1 A + \frac{4}{9} c_3 A^3\right][\cos(\omega_1 t) + \cos(\omega_2 t)] \tag{38}$$

注意，展开式中的二次因子构成了 DC 项的一部分，它加到输出偏置上。三次因子加大了基波项，但由于加大的倍数正比于幅值的立方，因此它的影响要比单单增加增益要大。一般地说，DC 的偏差来自于级数展开式中的偶次幂，而基波项则来自于奇次因子。

这里还有二次和三次谐波项，它们分别是由级数展开式中的二次和三次因子引起的：

$$\left[\frac{c_2 A^2}{2}\right][\cos(2\omega_1 t) + \cos(2\omega_2 t)] + \left[\frac{c_3 A^3}{4}\right][\cos(3\omega_1 t) + \cos(3\omega_2 t)] \tag{39}$$

一般地，第 n 次谐波来自 n 次因子。由于谐波失真的乘积具有比基波高得多的频率，通常在调谐放大器中被充分衰减，因此占支配地位的是其他非线性乘积。

二次项也造成了二次交调(IM)乘积项，就如在混频器中那样：

$$\left[\frac{c_2 A^2}{2}\right][\cos(\omega_1 + \omega_2)t + \cos(\omega_1 - \omega_2)t] \tag{40}$$

与谐波失真乘积项一样，如果 ω_1 和 ω_2 接近相等(如在这里假设的那样)，那么这些“和频”和“差频”项在窄带放大器中就会被有效地衰减掉。

最后，三次项造成了三阶交调乘积项：

$$\left(\frac{3c_3 A^3}{4}\right)[\cos(\omega_1 + 2\omega_2)t + \cos(\omega_1 - 2\omega_2)t + \cos(2\omega_1 + \omega_2)t + \cos(2\omega_1 - \omega_2)t] \tag{41}$$

注意这些乘积项随驱动幅值的立方而加大。一般地，一个 n 次 IM 乘积项的幅值正比于驱动幅值的 n 次幂。

那些和频的三阶 IM 项在调谐放大器中的重要性不大，这是因为它们一般在频带之外足够远处所以被显著地衰减。然而差频分量可以是很成问题的，因为如果 ω_1 和 ω_2 只差一个很小的数量(例如，在一个信号和一个相邻信道干扰信号的情形中)，那么这些差频就可能落在频带之内。正是由于这个原因三阶交调点是线性度的一个重要度量。

由上面一系列公式很容易计算出折算到输入端的三阶交调(IIP3)，即只要使 IM3 积项的幅值等于线性基波项的幅值：

$$|c_1 A| = \left|\frac{3}{4} c_3 A^3\right| \Rightarrow A^2 = \frac{4}{3}\left|\frac{c_1}{c_3}\right| \tag{42}$$

式中已假设了在表示基波输出幅值时只是稍微偏离线性。重要的是要强调交调是一个外推的值，因为由式(42)计算出的相应幅值几乎总是很大因此舍去级数三次项以后的各项会引起显著的误差。因此无论是在模拟还是在实验中，交调是在较小幅度输入时由观察到的外推趋势来估计的。在这样小的输入条件下，由高阶非线性贡献给基波的那些项是可以被忽略的。

因为由式(42)得到的是电压幅值的平方，所以除以输入电阻 R_s 的两倍就得到了通过外推使 IM3 等于基波项时的功率：

---

⑩ 这一推导多次用了以下三角恒等式：$\cos x \cos y = [\cos(x+y) + \cos(x-y)]/2$

$$\text{IIP3} = \frac{2}{3}\left|\frac{c_1}{c_3}\right|\frac{1}{R_s} \tag{43}$$

我们看到，IIP3 是正比于在偏置点求得的转移特性的一阶与三阶微分的比值。或等效地说，它是正比于小信号增益与那个增益的二阶导数的比(再一次，这是在相应的偏置点上求值的)。

图 13.10 总结了这些线性度的定义。图中，习惯上把输出功率画成两个(幅值相同的)不同频率的输入正弦波中每一个功率(它们的幅度是相等的)的函数，而不是它们和的函数。

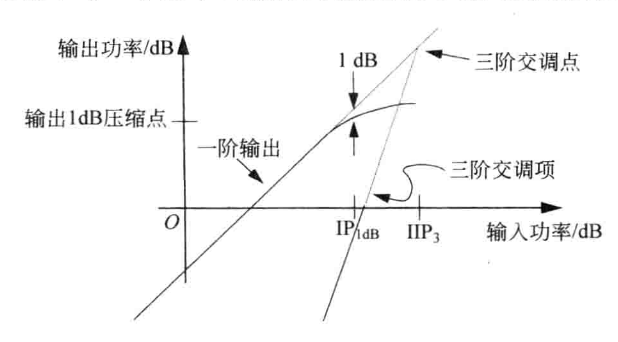


图 13.10　LNA 性能参数的图示

由于那些三阶乘积项随驱动幅值的立方而增加，因此当画在对数坐标上时它们的斜率是一阶输出乘积项的三倍，如图 13.10 所示。注意，在图中 1dB 压缩点发生在比 IIP3 低的输入功率处。这一个一般关系在实际的放大器中几乎总是这样(而且有较充分的余量)的。

在定义了线性度的度量后，我们现在考虑利用或不利用式(43)来估计 IIP3 的方法。

估计 IP3 的一些方法

求 IP3 的一种方法是通过瞬态模拟，即使两个幅值相等频率几近相同的正弦输入信号驱动这个放大器。当输入幅值改变时，将输出频谱中的三阶交调乘积项与基波项加以比较，然后计算相交点。

尽管在原理上很简单，但这一方法有几个明显的实际困难。首先，由于失真的乘积项可以比基波项小几个数量级，所以模拟器的数值噪声很容易在输出中占支配地位除非规定特别严格的容差[11]。一个密切相关的考虑是模拟所用时间步长必须间隔相同并足够小以免在输出谱中引起假象[12]。当满足这些条件时，模拟一般都会执行得非常慢并会生成很大的输出文件。纯频域的模拟器(例如谐波平衡工具)可以在短得多的时间内计算出 IP3，但比起时域模拟器如 Spice 来，当前还不能普遍得到这一工具。

式(43)采用功率级数两个系数的比提供了三阶交调的简单表达式，因此提示我们有另一种可适合手工计算的方法。尽管我们很少会被直接给予这些系数，但如果已经有了对传输特性的解析表达式，那么确定这些系数是一件很容易的事。即使没有这样的一个表达式，那么也有极为简单的步骤可以很容易地在“通常”的模拟器如 Spice 中实现，以便快速地估计 IP3。这一被我们称为三点法的技术，利用了这样一个事实，即已知在三个不同输入幅值时的增量增益就足以确定三个系数：c_1，c_2 和 c_3[13]。

---

⑪ 事实上，容差要比通常提供的“精确”解的默认选择严格得多。

⑫ 这一要求来自于实际模拟器采用的所有 FFT 算法所作的假设，即时间采样是均匀间隔的。

⑬ 这一方法是采用了真空管时代用来估计谐波失真的经典技术，稍加改动得到的。

为了推导这个三点法，我们从联系输入和输出的级数展开式开始：

$$i(V_{DC}+v) \approx c_0 + c_1 v + c_2 v^2 + c_3 v^3 \tag{44}$$

其增量增益（即跨导）是式(44)的导数：

$$g(v) \approx c_1 + 2c_2 v + 3c_3 v^2 \tag{45}$$

尽管在原理上任何三个不同的 v 值都可以，但特别方便的三个是 0，V 和 $-V$，这些电压可以看成是从 DC 偏置值的偏移。当选择这些值时，得到了以下相应增量增益的表达式：

$$g(0) \approx c_1 \tag{46}$$

$$g(V) \approx c_1 + 2c_2 V + 3c_3 V^2 \tag{47}$$

$$g(-V) \approx c_1 - 2c_2 V + 3c_3 V^2 \tag{48}$$

求解这些系数得到：

$$c_1 = g(0) \tag{49}$$

$$c_2 = \frac{g(V)-g(-V)}{4V} \tag{50}$$

$$c_3 = \frac{g(V)+g(-V)-2g(0)}{6V^2} \tag{51}$$

将这最后三个系数公式代入到式(43)中就得到了所希望的用三个增量增益表示的 IIP3 的表达式⑭：

$$\mathrm{IIP3} = \frac{4V^2}{R_s} \cdot \left| \frac{g(0)}{g(V)+g(-V)-2g(0)} \right| \tag{52}$$

用式(52)求 IIP3 比用瞬态模拟快得多，因为无论用模拟器还是人工手算，确定增量增益几乎没有什么太大的计算量。因此三点法在设计早期阶段在用于快速估计 IIP3 是特别有用的。这个方法在指导取得最大化的 IP3 是偏置点（或电路拓扑）选择上也是十分有用的。注意，从式(52)可以看出，如果能够被允许选择一个偏置点，或者是被允许确定一个电路，以使得在那个偏置点的小信号增益等于与这个偏置点偏离 $\pm$V 的偏置处增益平均值，那么 IIP3 会变得无穷大。比如，如果一个器件的小信号跨导与偏置点的变化关系如图 13.11 所示，那么偏置点 A 或 C 会使 IIP3 最大化。小结一下：如果最大化 IIP3 是设计的目标，那么应该将放大器偏置到一个中间位置，在该处小信号增益随输入幅度线性的变化⑮。

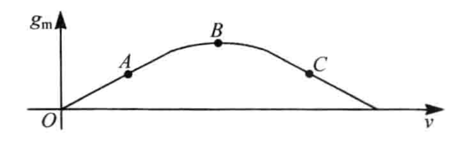


图 13.11　假想的放大器跨导(g_m 与 v)

13.7　赝空动态范围(SFDR)

至今我们已经认定了对允许的输入信号幅值的两个基本限制。噪声系数确定了下界，而失真则决定了上界。因此，放宽一点来说，放大器可接受的信号范围是下至噪声低限上至

⑭　在确定了用易于测量的增益表示的所有系数后，就很容易推导出对谐波和二阶 IM 失真的类似表达式。特别是这后一个量与直接变换的接收机有关。

⑮　对于 IP2 估计的一个二点法的相应推导揭示了放大器应该被偏置在一个值，该处的小信号增益的一阶导数是零，这是假定设计目标是最大化 IP2。一般来说，将放大器偏置至其小信号增益的 n 阶导数为零，可以最大化 IP($n+1$)。

某种线性度的限制。采用一个动态范围度量可以帮助设计者避免犯一个错误，即在改善一个参数（例如，噪声系数）的同时无意中使另一个参数变差。

通过称为赝空动态范围（spurious-free dynamic range，SFDR）的参数可使这一概念具有定量的基础。“赝”（spurious）这一词是指“不希望”“伪”的，并且常常被简写成“spur”[16]。在谈及 LNA 时，它通常是指三阶乘积项，但有时也指输出频谱其他不希望的分量。

为了理解采用 SFDR 作为动态范围的具体度量所基于的理由，把信号噪声比或信号失真比中较小的一个定义为一个更为一般性的度量，并且通过改变加到放大器上两个（正弦）输入的幅值然后估计这个度量。当输入幅值从零起增加时，一阶输出（即基波）最初具有负的信号噪声比但最终会超出噪声的最低限度。由于三阶失真依赖于输入幅值的立方，此时对任何实际的放大器，IM3 的乘积项将足够低于这一噪声低限。因此，当输入信号继续增加时，动态范围在一个时期内得到改善，因为此时所希望的输出增加了而不希望的输出（这里是噪声）仍然固定不变。然而三次 IM 项最终也会超出噪声的低限。当超出了这一输入电平后，动态范围就会下降，因为 IM3 项增长的速度是一阶输出增长速度的三倍（这是基于 dB 而言的）。

这个 SFDR 定义为当不希望的乘积项（这里是三阶 IM 功率）恰好等于噪声功率时相对应的输入幅值的信号噪声比，因而也就是在上面这一段中描述一个放大器所显示的最大动态范围。如图 13.12 所示。

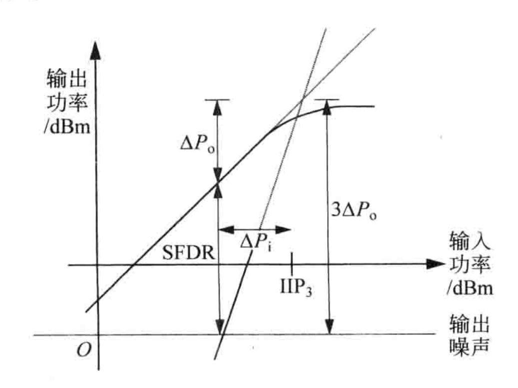


图 13.12　赝空动态范围（三阶）

为了在 SFDR 表达式中显式引入噪声系数和 IIP3，首先定义 N_{oi} 为用分贝表示的折算到输入端的噪声功率。由于采用 dB 坐标时三阶 IM 乘积项的斜率为 3，于是，当 IM3 的输出功率等于 N_o 时，输入功率比 IIP3 低的部分为

$$\Delta P_i = \frac{\mathrm{OIP3} - N_o}{3} \tag{53}$$

（同样，所有的功率都用分贝表示）。所以 SFDR 就是式（53）所对应的输出功率与 N_o 之间的差：

$$\mathrm{SFDR} = (\mathrm{OIP3} - \Delta P_o) - N_o \tag{54}$$

[16] 尽管“spurious”不是一个拉丁词，有时（而且是错误地）却用“spurii”来表示其复数；“spurs”是一个应该采用的复数形式。

因为三阶 IM 乘积项的斜率是 3，我们知道，

$$3\Delta P_o = \text{OIP3} - N_o \tag{55}$$

因此

$$\text{SFDR} = \frac{2}{3}[\text{OIP3} - N_o] = \frac{2}{3}[\text{IIP3} - N_{oi}] \tag{56}$$

可以观察到折算到输入端的噪声功率（这里用瓦特表示）就是噪声因子 F 乘以噪声功率 $kT\Delta f$。同时注意折算到输入端的量可以用在式(54)中，因为式中的两项（OIP3，N_o）都为同一个增益因子所缩小。

令人满意的是，SFDR 确实在一端为 IIP3 所限制而另一端为噪声的低限所限制，这正如在本节开始时我们定性说明的那样。之所以因子等于 2/3 是因为在定义这些限制时采用了特定的方式。

13.8 级联系统

一个级联系统的总的噪声指数与动态范围依赖于每一级的噪声指数、交调点及增益。之所以依赖于增益，是因为信号一旦被放大，之后的各级的噪声就不太重要了。因此，有一种趋向，就是接收机前几级的噪声特性几乎控制着整个系统的噪声系数。

单级的噪声系数如何组合以产生总的噪声指数会变得很复杂，这是由于典型的系统中用到不同的阻抗。为了为系统的噪声系数建立一个方程，让我们看一下图 13.13 中的方块图。其中 F_n 是一个噪声因子，G_n 是一个功率增益（具体来说是可获取增益，即负载匹配时得到的增益）。由于噪声因子依赖于信号源的电阻，因此为了保证计算的正确性，必须计算每级的与前级的输出阻抗相对应的噪声系数。这个问题在分立元件电路的设计中不是很常见，这是因为在那里阻抗大小总是标准化的，但是在集成电路实现中则需要谨慎对待。

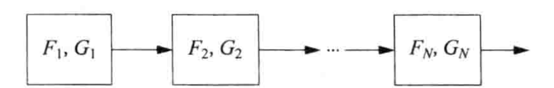


图 13.13　对噪声系数进行计算的级联系统

噪声因子可以用几种形式来表述，对我们现在处理的这个问题有一种表述特别有用：

$$F = \frac{R_s + R_e}{R_s} = 1 + N_e \tag{57}$$

其中，R_e 是一个（可能的）假想电阻，用来表示观察到的超过 R_s 所造成的噪声的那部分。因此 N_e 这个量是一个附加噪声功率的比，它等于 $F-1$。

把这个功率比折算到前一级的输入，可以简单地除以前级的可获取功率增益。将给定级的附加噪声贡献折算到系统的输入端，要除以在给定级和系统输入端之间总的可获取增益。总的噪声因子就是每级贡献的和，因而可以用下式表示：

$$F = 1 + F_1 - 1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \cdots + \frac{F_N - 1}{\prod_{n=1}^{N-1} G_n} \tag{58}$$

该式可以化简为

$$F = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \cdots + \frac{F_N - 1}{\prod_{n=1}^{N-1} G_n} \tag{59}$$

采用噪声温度，级联系统的总的噪声温度可以表示为

$$T_e = T_{e1} + \frac{T_{e2}}{G_{av1}} + \frac{T_{e3}}{G_{av1} G_{av2}} + \cdots \tag{60}$$

从最后这个表达式可以看出，为什么人们喜欢用噪声温度表征级联系统的噪声特性：这是因为总的噪声温度是所有折算到输入端的噪声温度的和。这样的一种说法并不适用于噪声因子。

很明显系统噪声系数的确是由最前面少数几个增益级的噪声特性起决定作用的。因此，在试图要得到好的噪声系数时，大部分的设计努力一般是集中在最前面的少数几级上。

级联系统的线性度

与系统的动态范围相关的另一个指标是交调（失真）点。尽管我们只讨论了三阶交调点，但是值得一提的是也有一些场合，在那里二阶交调点用作相关的线性度度量。一个众所周知的例子就是超外差结构的简并情形，即中频为零时。这样的直接变频接收器在近年来变得越来越普遍。

建立所需要的方程的一个困难在于发生某一级上的失真乘积项和它后一级上的失真乘积项的组合是与它们的相对相位有关的。因此，在单级交调和系统总交调之间没有简单而固定的关系。然而，我们可以估计出一个保守的（即最坏的）情况，就是把失真乘积项的幅度简单地相加。相对于用功率增益表示系统噪声指数来说，把增益表示成为电压比的形式对上述这种选择显得非常自然。图 13.14 描述了这种情况，其中 A_{vn} 是电压增益，$IIVM_n$ 是 M 阶输入交调电压。

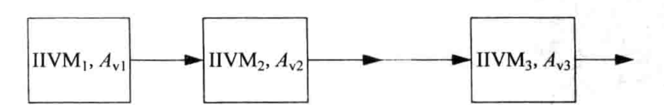


图 13.14　计算输入交调点的级联系统

为了方便推导，我们用 $V_{dM,n}$ 来表示当电压 V 加到第 n 级的输入端时，折算到输入端的 M 阶交调失真积的电压。注意到，由输入交调的定义，M 阶 IM 失真积可以写为

$$V_{dM} = \frac{V^M}{IIVM^{M-1}} \tag{61}$$

让我们来推导只有两级的一个特定的级联系统的三阶交调。折算到输入端的第一级三阶 IM 为

$$V_{d3,1} = \frac{V^3}{IIV3_1^2} \tag{62}$$

在第二级输出端的三阶 IM 电压 $V_{d3\ out,tot}$ 是两部分的和：一部分就是经过第二级放大的第一级的失真积，第二部分是第二级产生的失真积。这两部分相加得到下面的偏悲观点估计：

$$V_{d3,tot} = A_{v2} V_{d3,1} + V_{d3,2} \tag{63}$$

折算到输入端的三阶失真电压就是再除以总的增益 $A_{v1} A_{v2}$：

$$V_{d3\ in,tot} = \frac{A_{v2} V_{d3,1} + V_{d3,2}}{A_{v1} A_{v2}} \tag{64}$$

把式(62)代入式(64)，并注意到式(61)可用于来求式(64)中的 $V_{d3,2}$，如果 V 用 $A_{v1}V$ 来代替的话，这样就得到

$$\frac{1}{\mathrm{IIV3}_{\mathrm{tot}}^2} = \frac{A_{v1}^2}{\mathrm{IIV3}_1^2} + \frac{(A_{v1}A_{v2})^2}{\mathrm{IIV3}_2^2} \tag{65}$$

由最后一个方程可以看出，因为前级的放大作用，所以后级承担了较重的负担。我们也可以看到一个给定级的 IIV3 的倒数是先乘以到该级输入端的总增益后，再加以平方，然后与其他级加起来，这样得到的和就是折算到输入端的总的交调倒数的平方。

尽管式(65)只适用于两级的级联，但是它可以扩展到任意级的级联系统的情况：

$$\frac{1}{\mathrm{IIV3}_{\mathrm{tot}}^2} = \sum_{j=1}^{n}\left\{\frac{1}{\mathrm{IIV3}_j^2}\prod_{i=1}^{j}A_{vi}^2\right\} \tag{66}$$

级联的噪声系数与线性度的计算很快能变得枯燥乏味与臃长，特别是当为了确定与取得某个优化点而需要进行迭代计算时。只要花足够大的气力，总是能构造一个数据表来完成这一类的计算，而这也正是许多工程师多年来一直这么做的。幸运的是，大量的将这个过程自动化的工具现在是可以免费得到的。一个很受欢迎的工具是安捷伦(Agilent)公司的 AppCAD，这个软件可以产生如图 13.15 所示的例子。

	Preselector	LNA	Image Filter	Post amp	Mixer	IF Filter	IF amp
[illegible]	1.5	0.8	3.0	2.0	10	3.0	8
[illegible]	-1.5	15	-3.0	15	12	-3.0	60
[illegible]	100	15	100	10	20	100	15
[illegible]	0	0	0	0	0	0	0
[illegible]	0	0	0	0	0	0	0
[illegible]							
[illegible]	1.50	0.80	3.00	2.00	10.00	3.00	8.00
[illegible]	-1.50	15.00	-3.00	15.00	12.00	-3.00	60.00
[illegible]	-50.00	-51.50	-36.50	-39.50	-24.50	-12.50	-15.50
[illegible]	-51.50	-36.50	-39.50	-24.50	-12.50	-15.50	44.50
[illegible]	0.79	0.94	0.05	0.09	0.02	0.00	0.00
[illegible]	-0.21	-0.06	-0.04	-0.01	0.00	0.00	0.00
[illegible]	0.00	0.00	0.00	0.01	0.03	0.00	0.94

Enter System Parameters:

[illegible]	-50	dBm
[illegible]	25	℃
[illegible]	1	MHz
[illegible]	25	℃
[illegible]	10	dB
[illegible]	290	°K

System Analysis:

Gain =	94.50	dB
Noise Figure =	2.61	dB
Noise Temp =	238.45	°K
SNR =	61.37	dB
MDS =	-111.37	dBm
Sensitivity =	-101.37	dBm
Noise Floor =	-171.37	dBm/Hz

Input IP3 =	-19.72	dBm
Output IP3 =	74.78	dBm
Input IM level =	-110.55	dBm
Input IM level =	-60.55	dBC
Output IM level =	-16.05	dBm
Output IM level =	-60.55	dBC
SFDR =	61.10	dB

图 13.15　级联系统计算的例子(采用 AppCAD)

类似于这样的分析不仅仅对评估性能来说是非常有价值的，而且对确定在一个系统中最限制性能的电路级也是如此。作为一个方便的特性，AppCAD 凸显了与控制整个系统的噪声系数与线性度最为强烈的电路级相对应的那些电路单元。在图 13.15 的例子中，我们看到输入的前置选择滤波器与 LNA 对噪声系数有最大的影响(这是意料之中的)，而中频放大器的线性度则在很大程度上决定了整个系统的线性度(又一次，这也是意料之中的)。假若我们的目标是去找到系统性能在这两个度量(噪声与线性性)中任一个方面是差的话，我们或许会知道设计努力应该最有效地用在什么地方。

13.9 总结

我们已经看到一个电感简并的 LNA 同时取得了极佳的阻抗匹配，接近于最小的噪声系数与不错的增益。

我们也介绍了计算线性度的三点式方法，这个方法允许得到一个近似的，但是定量的对线性度的估量，这比起直接采用时域模拟器要快得多。尽管这个方法忽略了动态行为，对实际放大器的测量通常是揭示了与预计相比十分好的一致性。只要器件是工作在比 ω_T 低许多的频率下，通常总是可以期待得到不错的结果符合。

假若要求更好的线性度，或者是功耗，或者是增益需要牺牲以来换取改进了的线性度。比如说，可以通过改变偏置条件来降低输入的 Q 值，或者采用负反馈。将被噪声品质因子与失真品质因子的考虑而对信号幅度的限制结合起来，就得到了一个对放大器的最大动态范围的度量，这就是赝空动态范围。

最后，我们研究了计算一个系统级联的噪声系数与估计交调点的一些方法。

13.10 附录 A　双极型晶体管的噪声系数方程

我们在这里重复一下来求得最佳源电阻（包括“精确”与近似）的方程与相应的噪声因子：

$$R_S^2 = \frac{2|z_\pi|^2 g_m r_b + r_b^2 + |z_\pi|^2 + 2r_b \mathrm{Re}\{z_\pi\} + g_m^2 r_b^2 |z_\pi|^2/\beta_F}{1 + |z_\pi|^2 g_m^2/\beta_F} \tag{67}$$

$$R_S^2 \approx \frac{\left(\dfrac{\omega_T}{\omega}\right)^2 \left(\dfrac{2r_b}{g_m} + \dfrac{r_b^2}{\beta_F} + \dfrac{1}{g_m^2}\right) + r_b^2 + \dfrac{2r_b r_\pi}{(\omega/\omega_T)^2 \beta_F^2}}{1 + \dfrac{(\omega_T/\omega)^2}{\beta_F}} \tag{68}$$

$$F = 1 + \frac{r_b}{R_S} + \frac{(R_S + r_b)^2 + |z_\pi|^2 + 2(R_S + r_b)\mathrm{Re}\{z_\pi\}}{2R_S |z_\pi|^2 g_m} + \frac{(r_b + R_S)^2 g_m}{2\beta_F R_S} \tag{69}$$

下面这个对阻抗 z_π 的表达式或许是有帮助的：

$$|z_\pi| = r_\pi / \sqrt{(\omega r_\pi C_\pi)^2 + 1} \tag{70}$$

$$\mathrm{Re}\{z_\pi\} = r_\pi/(\omega r_\pi C_\pi)^2 + 1 \tag{71}$$

13.11 附录 B　场效应晶体管（FET）噪声参数

13.11.1 理论

MOSFET、JFET 及 MESFET 的基本噪声模型是由两个本征源组成的。我们一开始将忽略非本征的噪声源，例如那些与栅电极材料相关联的损耗引起的噪声，见图 13.16。

图 13.16　MOS 噪声模型

一个本征的噪声电流发生器是从源端连到漏端，其平均方值为

$$\overline{i_{nd}^2} = 4kT\gamma g_{d0}\Delta f \tag{72}$$

其中，γ 的理论值是 2/3，g_{d0} 是 V_{DS} 为零时的漏源电导。

还有一个栅电流噪声，是用一个从栅连到源的电流源来建模，其均方值为

$$\overline{i_{ng}^2} = 4kT\delta g_g \Delta f \tag{73}$$

其中

$$g_g = \omega^2 C_{gs}^2 / 5g_{d0} \tag{74}$$

δ 的理论值是 4/3。

更进一步说，栅噪声与漏噪声是互相关联的，其关联系数形式上定义为

$$c \stackrel{\text{def}}{=} \overline{i_{ng} \cdot i_{nd}^*} / \sqrt{\overline{i_{ng}^2} \cdot \overline{i_{nd}^2}} \tag{75}$$

长沟器件的 c 值理论上是 j0.395。为了简化推导起见，我们也忽略了 C_{gd}。尽管可达到的噪声系数几乎不受 C_{gd} 的影响，输入阻抗却可以是 C_{gd} 的强函数，因此当设计输入匹配网络时这个效应必须被考虑在内。

我们这里忽略噪声参数的推导，只是将结果总结在表 13.2 中，其中 $\alpha = g_m / g_{d0}$。

表 13.2　FET 两端口噪声参数小结

参　　数	表　达　式
G_c	~ 0
B_c	$\omega C_{gs}(1-\alpha\lvert c\rvert\sqrt{\delta/5\gamma})$
R_n	$\dfrac{\gamma g_{d0}}{g_m^2} = \dfrac{\gamma}{\alpha} \cdot \dfrac{1}{g_m}$
G_u	$\dfrac{\delta\omega^2 C_{gs}^2(1-\lvert c\rvert^2)}{5g_{d0}}$

使用这些参数，我们能够既决定使噪声系数最小化的源阻抗，又能决定最小的噪声系数本身：

$$B_{opt} = -B_c = -\omega C_{gs}(1 - \alpha \mid c \mid \sqrt{\delta/5\gamma}) \tag{76}$$

从式(76)可以看出，最优的源电纳本质上来讲是电感性的，除了其频率依赖关系与电感不同。因此，取得一个宽带的噪声匹配从根本上来说是困难的。

再继续推下去，最优的源电导为

$$G_{opt} = \sqrt{G_u/R_n + G_c^2} = \alpha\omega C_{gs} \sqrt{(\delta/5\gamma)(1 - \mid c \mid^2)} \tag{77}$$

而最小的噪声系数则由下式给出：

$$F_{min} = 1 + 2R_n[G_{opt} + G_c] \approx 1 + \frac{2}{\sqrt{5}} \frac{\omega}{\omega_T} \sqrt{\gamma\delta(1 - \mid c \mid^2)} \tag{78}$$

在上面这个表达式中，如果 ω_T 是简单的处理成为 g_m 对 C_{gs} 的比，那么这个近似式就变成完全准确的了。注意，如果没有栅电流噪声（即，假若 δ 为零），那么最小噪声系数会是 0dB。这个不切实际的预测本身就足够说明栅噪声的的确确是存在的。同时也要注意，原则上增加漏与栅电流噪声的关联度会改进噪声系数，然而为了使这个效应能引起噪声系数的显著下降，关联系数需要不切实际地接近于单位 1。

上面这些推导与讨论忽略了几个重要的因素。一个是与栅电极材料的非超导性（一句玩笑话，指电阻率非零，译者注）相关联的热噪声。尽管通过制作者的仔细的对器件的版图设计能使得这个值比较小，这个因素经常会大大提高最小的噪声系数。我们也忽略了与衬

底相关的热噪声。在微波频率下，对典型的 MOSFET 管，这个成分的确可以被忽略，但对于任何一个特定的情况，那里这个噪声成分可能有关，我们必须验证它能否被忽略。

13.11.2　实际考虑

为准确计算噪声所需的模型参数并不总是已知的。假若针对低噪声的设计还是要进行，必须开发一个替代的方法来得到这些模型参数。采用一个纯经验的方法可以完全不依赖理论假设，但是其实际的实现则要求一套恰当的，且有限数量的实验设计。幸运的是，两端口噪声系数维数本身就建议一个得到适合于在大范围的条件下使用的数据的简约方法。这个方法也适用于其他模拟模型的相关噪声系数的提取。

我们已经注意到两端口网络的噪声参数或者是阻抗或者是导纳。尽管它们不一定代表那些可以直接测量的物理量，它们还是服从同样的缩放规律。比如说，假设我们已完整地表征了具有某个单位尺寸的一个 MOSFET。对于固定的偏置电压(意味着固定的电流密度)，两个这样的器件并联起来，其阻抗参数一定是等于单个器件参数的一半，同时对导纳参数则是加倍。没有量纲的参数值则没有变化。把这个说法一般化，一个沟道宽带为 W 的器件其噪声参数为

$$G_c = \frac{W}{W_0}G_{c0} \tag{79}$$

$$B_c = \frac{W}{W_0}B_{c0} \tag{80}$$

$$G_u = \frac{W}{W_0}G_{u0} \tag{81}$$

$$R_n = \frac{W_0}{W}R_{n0} \tag{82}$$

这里，下标"0"表示对应于一个宽带为 W_0 的单位器件的一个参数。噪声因子 F 是一个没有量纲的量，因此是独立于宽带的(尽管产生这个给定的 F 的源导纳随宽带缩放)。

这些缩放定律意味着我们只需对一个单一的器件尺寸(我们把这种器件称为单位单元)进行噪声参数的表征。尽管这个表征必须包括在一个范围内的偏置的变动，它不必对器件的宽度进行扫描这一特性就很明显有其价值。比如说，所需的参数可以从噪声系数的等值线上提取。基于这些缩放关系的外推会是准确的，只要我们小心地选择一个足够大的参考器件以使得任何测试配件与版图的寄生参数可以忽略，以及只要设计者在设计时仅采用这个单位单元的整数倍。后面这个做法是在模拟电路设计中为了达到良好的匹配所公认的做法。

第14章 噪声系数测量

14.1 引言

性能一般的放大器最重要的性能指标之一是噪声系数NF,或噪声因子F。这两个参数在文献中互换使用,但是采用如下固定的习惯用法：用NF来表示噪声系数,并且定义它为$10\log F$。对于这两个参数的使用不那么严格区分(这反映了惯常的用法),但是上下文应该能告诉我们是否在用以分贝为单位的那个参数。

现在使用的噪声因子的定义是由贝尔实验室的Harald Friis ①最早正式提出的。作为其核心,这个定义涉及到信噪比(SNR)：

$$F \overset{\text{def}}{=} \frac{\text{SNR}_\text{i}}{\text{SNR}_\text{o}} \tag{1}$$

这个定义表明F是一个放大器藉以来降低输入信号信噪比的因子。因为这样的定义,这个因子绝不会小于1。这个定义看起来是如此简单且直截了当,其实在其中隐含着许多细节之处,很快我们就会认识到这实际上是给出了一个不完整的定义。噪声系数的精确测量依赖于对这些细节的充分了解,以及懂得如何来确定和纠正测量误差的来源。如我们将很快会看到的,自动噪声系数测试仪并没有消除对有经验、有知识的操作人员的需求。就如一种说法所提到的:“自动化的仪器仅仅让你在单位时间内犯更多错误”。这一章的目的是减少错误答案的发生率。

14.2 基本定义和噪声测量理论

一个重要的细节是测量噪声系数测量时的温度。具体来说,信号源的温度对噪声系数有显著的影响。从直观上来看,这个温度的影响可以被理解如下：待测器件(DUT)产生的内部噪声是独立于信号源温度的。如果后者是很低的,信号噪声也将相应的低,因而DUT增加的噪声那部分将有相对更大的作用。这样,测量得到的噪声系数会比信号源更热(从而也有更大的噪声)时所得到的更大。正是由于这个敏感性,一个有意义的噪声系数的比较要求测量在一个标准温度下进行。Friis提出了一个参考温度,记做T_0,是290K(62华氏度或17℃),这个温度比大多数实验室的内部温度要低不少。对于这样一个选择的一个常常被引用的理由是这个温度近似等于地面无线通信使用的天线通常所遇到的温度。然而,这个选择的一个更可能的动机仅是因为这样kT_0会等于4.00×10^{-21}J,而在计算尺时代,这样一个整数值无疑极具吸引力。尤其是对于像Friis这样一个极其务实的人。

对于标准测试条件的最终规定是由无线电工程师学会(Institute of Radio Engineers,

① “Noise Figures of Radio Receivers”(无线电接收器的噪声系数),Proc. IRE, July 1944, pp. 419-422.

IRE,IEEE 的先驱)的委员会作出的：噪声系数的测量采用一个信号源,其可用噪声功率与温度处在 290 K 的输入端终端(电阻)产生的可用噪声功率一样。回顾一下,可用功率是定义为能够被输送到一个(共轭)匹配的负载的功率。因此,即使信号源实际上并没有驱动一个匹配的负载,这个功率依然是可取得的。可用功率就是如其字面上所指的：一个潜在的与实际负载无关的可取得功率。这个标准化委员会接受了 Friis 的有关将噪声系数建立在可用功率的基础上的建议,因为这个参数可以和一个热噪声源,例如电阻的温度直接关联起来。这个定义常被混淆,并且会导致严重的错误,这一点在这章后面将得以澄清。

第二个考虑是要确定输入和输出的信噪比绝不是一件轻而易举的事[②]。因为噪声系数是 DUT 本身的本征属性(我们假设器件是线性的,否则的话噪声系数就不能唯一地确定),因此与如何驱动 DUT 无关,应该可以设计一个不需要用外部信号源的测量方法。要这样做,注意到下面这一点是有帮助的：出现在 DUT 的输出端的可用噪声功率来自两部分贡献。一是被放大的可用信号源噪声功率(噪声源的温度是 $T_0 = 290$ K),其值为

$$N_{os} = kT_0 BG_{av} \tag{2}$$

其中,B 是噪声的(砖墙)带宽,G_{av}是 DUT 的可用功率增益。

输出噪声的另一个成分是为 DUT 自己添加的噪声。我们称该噪声贡献为 N_a。因此总的可用输出噪声功率为

$$N_1 = kT_0 BG_{av} + N_a \tag{3}$$

现在让我们重新访问与修正一下公式(1)给出的噪声系数定义：

$$F \overset{\text{def}}{=} \frac{\text{SNR}_i}{\text{SNR}_o} = \frac{S_i/N_i}{S_o/N_o} \tag{4}$$

如前所述,这个量必须在 290K 时求值。

设所有的量都为可用功率,输出信号 S_o 和输入信号 S_i 之比是可用增益 G_{av}。有用输入噪声功率就是 kT_0B,而可用输出噪声功率是 N_1,如式(3)所定义的。所以可以写为

$$F = \frac{S_i/N_i}{S_o/N_o} = \frac{1}{G_{av}}\left(\frac{N_a}{N_i}\right) = \frac{1}{G_{av}}\left(\frac{N_1}{N_i}\right) = \frac{N_1}{N_{os}} = \frac{kT_0 BG_{av} + N_a}{kT_0 BG_{av}} \tag{5}$$

最右面的那个表达式：

$$F = \frac{kT_0 BG_{av} + N_a}{kT_0 BG_{av}} \tag{6}$$

就是被 IRE 正式采用的定义[③]。作为一个测量的基础,初看起来这个公式比式(1)更为吸引人,因为它不包含任何与外部输入输出信号相关的项。采用式(6),噪声系数的测量就简化成了噪声、可用增益和带宽的测量。遗憾的是,直接基于该公式的测量仍然存在许多严重的实际困难。特别是不容易测量具有高精度的可用噪声带宽和可用增益的乘积 BG_{av}。在对 14.6 节将要描述的各种噪声测量方法进行比较之后,就可能充分领会到这个实验上的困难。

噪声系数测量的另一种已经在商用仪器,如 HP8970A,实现了的替代方法是巧妙地避开了直接测量增益-带宽测量的需要,代之以测量在两个不同温度下的噪声比。一般而言,只要对一个量的量纲考虑允许,把测量绝对的量变成测量这些量的比值总是有好处的。幸

② 第三个细微之处产生于含有频率变换单元的系统中,例如混频器。我们将在后续的 14.8 节中讨论这个问题。

③ 参见 *Proc. IRE.* v. 51, no. 3, March 1963, pp. 434-442。

运的是，噪声系数是一个没有量纲的量，所以一个纯比例的测量是可能的。增益-带宽乘积是有量纲的，因此从根本上说不应该一定要直接测量这个量。

比例量测技术的基础是基于这样一个事实：增加噪声源的温度造成了在输出噪声中因为这个外加噪声源而引起的分量，但并不改变由 DUT 引起的噪声。如果噪声源温度变化的比例是准确已知的，那么测量在噪声源热与冷的条件下的输出噪声功率会允许我们求解因 DUT 增添的噪声，这样就可以计算出噪声系数。

图 14.1 画出了输出噪声功率和噪声信号源温度的函数关系，诠释了这样的一个比例测量方法是如何解决了我们的问题的。将该图的特征与式(6)相比，可以注意到斜率和 y 轴上的截距告诉了我们计算 F 所需要的一切信息：

$$F=\frac{kT_0BG_{av}+N_a}{kT_0BG_{av}}=1+\frac{N_a}{kT_0BG_{av}}=1+\frac{y-\text{intercept}}{(T_0)(\text{slope})} \tag{7}$$

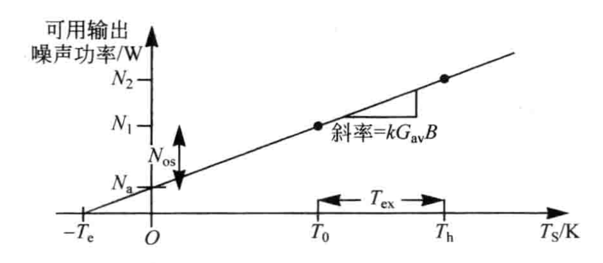


图 14.1　输出噪声功率与噪声信号源温度的关系

显然，测量增益和带宽的需要不再存在了，这是因为有两个点就能确定一条直线。但是尽管这个观察的结果似乎隐含着可以用一个十分简单的数据处理方法，工程师们还是设计了大量的不同方法来通过两个点处的噪声数据来确定噪声系数。只是我们需记住，在这些看起来复杂的表面之下是极其简单的为图 14.1 所示的几何图像。

如果我们在一个噪声源温度为 T_h 上进行噪声功率测量，T_h 比参考温度高 T_{ex}，那么有效输出噪声功率变成：

$$N_2=kBG_{av}T_h+N_a=kBG_{av}(T_0+T_{ex})+N_a \tag{8}$$

再把与 T_0 时(式(3))的热测量结合起来后，通过简单的代数运算就可以得到噪声系数的如下表达方式：

$$F=\frac{T_{ex}/T_0}{N_2/N_1-1} \tag{9}$$

在文献中，这个比例 N_2/N_1 经常被称为“Y 因子”(为什么呢？因为它在 X 后面，……)。图 14.1 所示为一个等于参考温度 T_0 的冷温度，但是需要清楚的是，任何除了 T_h 的温度都可用于计算直线的斜率和截距。更一般地说，如果冷温度 T_c 不是 T_0 的话，那么分子要发生变化，而噪声系数则变成：

$$F=\frac{T_{ex}/T_0-Y(T_c/T_0-1)}{Y-1} \tag{10}$$

比例 T_{ex}/T_0 是噪声源的一个特性，并且是个(几乎)由制造厂商提供的信息。之所以用“几乎”这个修饰词是因为制造厂商实际上定义了一个稍微有点不同的量，被称为过剩噪声率(ENR)，其定义为传递给一个 50Ω 负载(或偶尔是一些其他标准阻抗)的噪声功率比例。然

而，比例 T_{ex}/T_0 来自于对有效功率的考虑（Y 因子也是）。只有在噪声源正好具有精确的 50Ω 阻抗的特定情况下，这两个比例是等价的。尽管制造者已经尽了最大的努力，这个条件实际中并不可能完全得到满足，所以用 ENR 代替 T_{ex}/T_0 是一个（但一般较小）潜在误差来源。然而，因为确定 ENR 要容易得多，这就是为什么 NBS（国家计量局，the National Bureau of Standards，现在称为 National Institute for Standards and Technology，NIST）决定这样做的，因而也就是制造者所测量和报告的。

在“早先的日子”里，实际的热和冷电源通常采用 77K（液氮沸点）、373K（水沸点，尽管电阻是被电加热到这个温度，而不是浸入水浴）的电阻。显然，温度差越大，对于给定量级的功率测量中的不确定性，我们可以更加准确地计算出斜率和截距。对于热温度一侧的一个限制是准确地确定或控制温度。并且温度越高，材料特性的一些其他问题就会越显著（例如，熔化）。

现在，常用到噪声二极管[④]（参见第 9 章），它可以产生一个异常热的热源噪声（例如 10 000K，比任何已知金属的熔点都高），但是却保持在室温。同样地，二极管也可以产生冷参考源，只需把它关断，导致一个内部的电阻匹配网络来提供一个和环境温度一样的有效噪声功率。参见图 14.2，其中射频扼流器（RF choke，RFC）就是个电感，其感值大到足以在所有有关的频率下都可视为开路。

图 14.2　典型的噪声二极管

这种噪声源的一个缺点是：不像真正的热电阻和冷电阻，这种二极管并不是根本的标准；它们的热噪声不能由基本原理得以计算。因为 ENR 只有非常准确地被知道才有用，它通常是被追溯到一个由国家实验室（如 NIST）保存的噪声基准（这个基准是一个被加热的或冷却的物理电阻）。这种要求的可追溯性部分说明了为什么噪声二极管具有相对较高的成本[⑤]。

14.3　噪声温度

噪声温度 T_e 在某些场合下可以取代噪声系数作为一个替代的品质因子。如我们在图 14.1 所见到的，噪声温度是噪声功率相对于温度的曲线的延长线在温度轴的截距（再取负值）。通过将噪声功率曲线向右平移一个等于噪声温度的温度，可以直观地看出这个噪声温度的意义，见图 14.3。

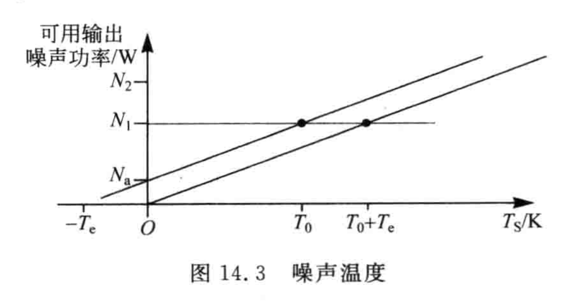


图 14.3　噪声温度

④　参见，例如 HP Journal（HP 期刊），April 1983，p. 26。

⑤　然而，对于这个高成本的一个更为重要的解释仅仅是市场可以承受到什么程度。

这个平移过的曲线是一个具有与原始的放大器同样斜率(=有用增益带宽积，再乘以 k 倍)的无噪声放大器的(因为在零噪声源温度时输出的噪声是零)。能够看出，如果噪声源被加热到温度 T_0+T_e，这个无噪声放大器会产生一个等于原来的放大器的可用输出噪声功率。由于更热的噪声源而引起的可用噪声功率的增加与因原来的 DUT 增加的可用噪声(N_a)严格相等：

$$N_a = kT_eBG_{av} \tag{11}$$

有几个原因使得噪声温度被用来表征卫星通信系统。一是在空中的物体通常并不处于一个接近于 290K 的有效温度，所以选择 290K 这样一个参考温度缺乏足够的基于物理的理由。另一点是空间通信系统一般都具有超常低的噪声系数，而噪声温度对于非常低的噪声系数值有更高的分辨率度量。表 14.1 在一个通常被认为是噪声非常低的范围内对噪声系数、噪声因子和噪声温度进行了比较。

表 14.1 噪声系数、噪声因子和噪声温度的比较

NF/dB	F	T_e/K
0.5	1.122	35.4
0.6	1.148	43.0
0.7	1.175	50.7
0.8	1.202	58.7
0.9	1.230	66.8
1.0	1.259	75.1
1.1	1.288	83.6
1.2	1.318	92.3

有时注意到这些结论是很有帮助的：在噪声系数非常低的情形下(比如，低于 1dB)，用分贝表示的噪声系数约等于噪声温度除以 70～75。换句话说，dB 的每十分之一对应大约 7.0～7.5K。

为了把噪声温度和噪声因子联系起来，让我们回到噪声系数的 IRE 正式定义：

$$F \overset{\text{def}}{=} \frac{N_1}{N_{os}} = \frac{kT_0BG_{av}+N_a}{kT_0BG_{av}} \tag{12}$$

把式(11)代入消去 N_a 得到

$$F = \frac{kT_0BG_{av}+N_a}{kT_0BG_{av}} = \frac{kT_0BG_{av}+kT_eBG_{av}}{kT_0BG_{av}} \tag{13}$$

这个式子简化成

$$F = 1+T_e/290 \tag{14}$$

如果由 DUT 增加的噪声等于噪声源噪声功率，那么噪声系数将是 3dB，对应噪声温度为 290K。许多有效噪声温度在 100K 以下的 LNA 是市场上能买到的。

噪声温度可以结合式(14)和式(9)来间接发现，或者直接从图 14.1 的冷和热的噪声测量中得到。采用后者的方法，我们可写出：

$$N_1 = kT_cBG_{av}+N_a = k(T_e+T_c)BG_{av} \tag{15}$$

$$N_2 = kT_hBG_{av}+N_a = k(T_e+T_h)BG_{av} \tag{16}$$

所以有

$$Y = \frac{N_2}{N_1} = \frac{k(T_e + T_h)BG_{av}}{k(T_e + T_c)BG_{av}} \tag{17}$$

对 T_e 求解得到

$$T_e = \frac{T_h - YT_c}{Y - 1} \tag{18}$$

使用噪声温度的最常见的理由是在计算级联噪声系数时，$F-1$ 这个量经常反复出现，这一点我们在14.4节中将会看到。重新调整式(14)，显然噪声温度 T_e 是和 $F-1$ 成正比的，所以使用噪声温度可使这样的运算大为简化。

因为噪声系数和噪声温度每一个都完全包含另一方的信息(例如，恰如公式(14)所示)，你可以使用任何一个。选用哪个在很大程度上取决于使用者的文化背景和便利。

点噪声系数

许多情况下，人们对一个放大器的噪声性能作为频率的函数感兴趣。在那些情况下，测量的带宽是限制在某些给定的值(例如 HP8970 中是 4MHz)，然后在那个带宽中的噪声系数被报告为在某一特定频率下的值。因为这个参数是在一个特定频率点周围的窄带中测量的噪声系数，它被称为点噪声系数。文献中给出的噪声系数最经常的是点噪声系数。

14.4 级联系统噪声系数的 Friis 公式

对一个级联系统的噪声系数计算经常会是用错误的方式被执行。这再一次说明，问题是出在没有注意到某些细节之处。一个困难是，不像增益单个的噪声系数并不是以简单的方式结合起来以得到总的级联的噪声系数。另一点是系统的每一级会看的不同的电源阻抗，而噪声系数的计算是与该阻抗有关的。为了详细理解这些以及其他一些问题，我们现在推导出对于级联噪声系数的正确公式——称为 Friis 公式。

考虑一个带有噪声的的系统，其本身为另一个带有噪声的系统驱动，如图14.4所示。这个第一级在源电阻为 R_s 时测试得到的噪声因子 F_1 和可用功率增益 G_1。第二级则在将前一阶的输出电阻当作源电阻时测量得到可用功率增益 G_2 和噪声因子 F_2。如果有更多的电路级，每一级的可用功率增益和噪声系数都是通过将前一级的输出阻抗当做这一级的源电阻来确定的。常见的一个错误是将 R_s 当做所有级的源阻抗，但是这种选择只在输出阻抗正好是 R_s 时是正确的。在一个由分立模块构成的级联系统中这个要求才可能得到满足，但重要的是不要将那个通常情况是对的情形不恰当地一般化。

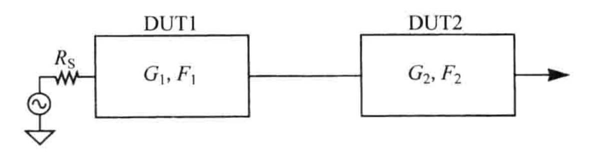


图14.4 级联系统

推导 Friis 公式最容易的方法是采用噪声温度的概念。因为被每一个 DUT 增加的可用输出噪声功率是 kT_eBG_{av}，在第一个 DUT 的输出端的可用噪声功率是

$$N_{o1} = kT_sBG_{av1} + kT_{e1}BG_{av1} = k(T_s + T_{e1})BG_{av1} \tag{19}$$

第二级接收了该噪声，将其放大，并且加上另一个它自己贡献的 kT_eBG_{av}：

$$N_{o2}=k(T_s+T_{e1})BG_{av1}G_{av2}+kT_{e2}BG_{av2} \tag{20}$$

我们也可以把整个系统看做一个可用增益为 $G_{av1}G_{av2}$ 的放大器，由一个电源 R_s 驱动。因此可以写出：

$$N_{o2}=k(T_s+T_{e12})BG_{av1}G_{av2} \tag{21}$$

其中，T_{e12} 是级联电路的总噪声温度。结合式(20)和式(21)得到：

$$T_{e12}=T_{e1}+\frac{T_{e2}}{G_{av1}} \tag{22}$$

因此，总噪声温度就是第一级的噪声温度加上第二级的输入参考噪声温度。这个公式反映了这样一个理解：第一级提供的信号增幅可以减小后续级的噪声效应。显然，式(22)可以扩展成任意阶数，这样就得到 Friis 级联噪声系数公式的一种形式：

$$T_{e12}=T_{e1}+\frac{T_{e2}}{G_{av1}}+\frac{T_{e3}}{G_{av1}G_{av2}}+\cdots \tag{23}$$

另一种以噪声因子表示的公式可通过式(14)将噪声温度和噪声因子相关联而推导出来：

$$F_{12}=F_1+\frac{F_2-1}{G_{av1}}+\frac{F_3-1}{G_{av1}G_{av2}}+\cdots \tag{24}$$

通过对最后两个等式的观察，我们看到采用级联噪声温度的表示法在某种程度上是更简单些(没有那个“－1”项来使公式看上去拥挤)。为第 n 级贡献的噪声温度可以通过除以前面($n-1$)阶的可用增益乘积来简单地计算得到。由于这个原因，对于级联系统，噪声温度公式常常更受欢迎。

14.5 噪声测量

从 Friis 公式我们可以看到，如果一个放大器有好的噪声系数但是低的增益，对后续级的噪声抑制就很差。遗憾的是，经典的噪声优化设计方法有时会导致一个正是含有这种特性组合的所谓“优化了的”放大器设计。因为放大器的噪声系数和增益大体来说都是重要的，另一个称为噪声度量的品质因数有时候被用来指导工程师进行综合平衡的设计。它的正式定义最初看上去是以有点奇怪的方式将这两个量结合起来：

$$M\stackrel{\text{def}}{=}\frac{F-1}{1-1/G_{av}} \tag{25}$$

但是当我们仔细看一下针对无穷级级联放大器的一个专门情形 Friis 公式后，就会清楚这个定义产生的逻辑性：

$$F_{tot}=F+\frac{F-1}{G_{av}}+\frac{F-1}{G_{av}^2}+\cdots \tag{26}$$

此式最终可简化成

$$F_{tot}=1+\frac{F-1}{1-1/G_{av}}=1+M \tag{27}$$

因此这个噪声度量的定义实际上是无穷级联的归一化了的噪声温度：

$$T_{e,tot}=(F_{tot}-1)T_0=MT_0\Rightarrow M=T_{e,tot}/T_0 \tag{28}$$

我们看到，对于好的噪声性能，希望噪声度量比器件自身的归一化了的噪声温度不大多少。

这里要提醒读者小心的是，噪声度量在一些文献中被定义成 F_{tot} 而不是 $F_{tot}-1$。一定

要搞清楚所采用的是哪一种定义，因为这个差异对于低噪声系统可能会引起相当大的误差。最后，要注意这个噪声度量的定义和负阻器件的噪声度量，如 Gunn 和隧道二极管（见第 9 章），没有什么特别的关系。

14.6 典型的噪声系数测量

在已经推导出噪声系数的多种表达式之后，我们现在要来看看实现实际测量的几种不同方法。就与以前一样，我们从简短地介绍历史开始，这部分是出于好玩，但是部分也是因为很久以前被使用的方法也是今天业余爱好者们可以经济地实现的方法。

14.6.1 过去的(好?)日子

从图 14.1 我们看到测量噪声系数和确定噪声功率一噪声源温度线的等式是等价的。测量线上的两点是足够的，但知道一点和线的斜率也是一样够了。前者是为现代所采用的方法，但后者值得在这里讨论。尽管这个方法提出了不容易对付的在实验上的挑战，但所需的仪器是大部分 RF 爱好者们力所能及的，所以关于该技术的描述值得在这里包括在内。

在开发出经校准的热、冷噪声源技术之前，仅有在室温下可用的噪声源。在那种限制下，人们只能在那个温度下（可能并不能准确地知道，并且不好控制）确定可用的输出噪声。所以我们立即就看到了一个误差来源：噪声源或许会在高于 290K 的温度。即使这样，这个误差源一般也不是主要的。

这个方法中有难度的部分是如何准确确定这条线的斜率，$kG_{av}B$。玻尔兹曼常数是很确定的，但是测量可用功率增益与噪声带宽（这通常不等于－3dB 带宽）的乘积是十分困难的。进行这种测量的实验装置的搭设在原则上是很直截了当的，但具体实现起来就困难了。

为了测量 $G_{av}B$，简单地把一个信号发生器连接到 DUT，然后扫描频率来画出功率增益对频率的曲线[⑥]。如图 14.5 所示。在大部分情况下，没有采取措施来确保共轭匹配（因为在每一个测试频率下都这样做是十分费功夫的），因此测量得到的功率增益和可用增益是不同的，这就导致了更多的潜在误差。功率频率响应曲线是积分得到的（比如，通过画图，或者通过测量－3dB 带宽并乘以某个在 1～1.57 间的拟合系数）来得到 GB 之积。

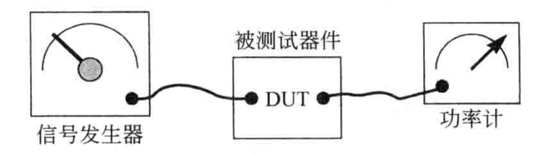


图 14.5 噪声系数测量的信号发生器方法

为了完成这个实验，输出功率 N_1 是在噪声源（比如通过一个简单电阻 R_s 来实现）连到输入端的条件学来测量的，则噪声因数为

$$F \stackrel{\text{def}}{=} \frac{N_1}{N_{os}} = \frac{N_1}{kT_0BG_{av}} \tag{29}$$

⑥ 如果信号发生器的输出在频带内不是常数，就需要测量其输出来进行适当的增益计算。进行这一过程是很乏味的（所以很容易被诱惑来省略这个步骤），但不这样做就构成理论一个常见的误差源。

这个测量方法只需要简单的仪器：一个信号发生器、一个已校准的功率计(或者示波器，见 14.10 节的通过目测来估计噪声的一些方法)和一个电阻(这可以简单地通过将信号发生器关掉来提供)。从测得的频响算出增益带宽积是十分费力气的，但是如果你不想在买一部汽车和买一个自动噪声系数计之间做选择，这个传统的方法就是最好的选择。话是这么说，用这个方法将噪声系数测量的不确定性降到 1～2dB 以下是很困难的，所以，实际上来说，用这种技术来表征噪声非常低的放大器一般是不可能的。

另一点是每个频率点的测量时间都挺长，所以要对改善噪声系数的方法的有效性进行实时验证是十分不方便的。采用这种信号发生器的方法需要很大的耐心。

但是存在着一种情形(至少，这是作者能想到的唯一的一种情况)，其中信号发生器方法可能实际上是有优势的。想象一个精确测量噪声很大的系统的噪声系数的问题。特别是，假设这个 DUT 有那么大的噪声，以至于噪声温度大大超过了参考温度。在这种情况下，热和冷的条件下的输出噪声功率可能很相近，导致 Y 因子接近于 1。因为热-冷测量法的噪声因子公式的分母中含有 $Y-1$ 项(见式(18))，当 Y 接近 1 时，测量可能对 Y 中出现的小的误差十分敏感。另一方面，对信号发生器法并不存在这个敏感性的问题，因为它不采用从测量两个温度的噪声导出斜率来。该方法中也没有减法。这样我们可以说：对于低噪声放大器，热一冷法更好，但是对于噪声极大的系统，信号发生器法可能更好。作者承认这个最后的情况可能是牵强的，因为在测量高噪声系数时，很少需要这样高的精度。

14.6.2 在当今这个时代

经校准了的热噪声源的存在大大便利了噪声系数的精确测量。如先前提到的，早期的噪声源采用了实在的电阻，加热或冷却到容易确定或控制的温度，如水和液氮的沸点。一个在水的沸点上的热噪声源对于在家里的实验者是完全可行的(只需小心不要烫到自己或引发火灾)并且十分准确——如果水是相当纯的并且修正了沸点随海拔高度的变化。然而，这类热源不够像商用仪器使用的那么热，所以精确性再次打了折扣。

加热了的电阻的一个替代方案是利用真空二极管的散粒噪声。当在有限温度的环境操作时，这种器件展现出的散粒噪声幅度可以追溯到第一性原理[7]。例如 5722，这是专门为这个目的设计的，这类温度受限的二极管有 $2qI$ 这么大的均方噪声电流密度，并且很容易产生几个(例如 5)分贝的 ENR。

许多商用的冷负载在 77 K 操作，但没多少业余爱好者会在家里有个充满液氮的杜瓦瓶[8]。或许对于业余实验者们的一个更实际的选择是用室温冷电源，但是准确测量需要知道实际的室温(尽管如此，你的测量也不可能那么精确)。采用冰水混合物来提供 273K 的冷温度，可以对测量精度有所改进。如果你有(不可饮的)乙醇和干冰，这两种物质达到平衡的混合物一般具有约 195 K 的温度(−78℃)。然而，酒精是可燃的，所以，如果确实选择了这种混合物，那么就要确保遵守所有适当的安全措施(尤其要使酒精远离任何加热噪声源的热源)。

---

⑦ 当工作在空间电荷受限的条件下时，会观察到不完全是散粒噪声的情形，这就使得计算变得复杂了。为此，这种二极管工作于阴极温度偏低且电流密度偏高的状态下，来保证温度受限的行为。

⑧ 假若你正好有液氮，那么一个对准一桶液氮的波导据说是实现 77K 冷源的好方法，

一旦你有了一个热源和冷源，有几种测量方式可供选择。一个方法叫做“Y 因子法”，使用这个方法的一些原因很快会变得清楚，它避免了对精确功率计的需求，代之以一个更易实现的经过校准的可调衰减器，见图 14.6。在这一类仪器中的衰减器经常就是一个工作在截止频率以下的波导，因为它容易做到使衰减量和机械参数相关联。

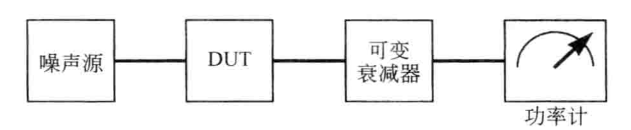


图 14.6　Y 因子测量技术(简化了的)

这个测量技术基于这样一个事实，即热源和冷源的输出功率比(= Y)加上已知热源和冷源的温度，就足够计算噪声系数了。要用这种方法做测量，先把衰减系数设成 1，连接冷负载，记录功率计上的输出功率示数。读数的绝对值是完全不重要的。然后连接热负载，并把衰减器调到使输出功率与刚才所示数相同。因为衰减因子就是将 N_1 减小到功率 N_2，衰减因子就精确地等于 Y。噪声因子因而就可以从式(10)计算出：

$$F=\frac{T_{ex}/T_0-Y(T_c/T_0-1)}{Y-1} \tag{30}$$

所能得到的精度依赖于 Y 因子确定的精度，以及有关热源和冷源的温度。如果精心控制所有的误差源的话，这种方法能实现商用仪表(假设噪声源的温度相同)的精度。所花的代价是每次测量的时间会较长。

实际的 Y 因子法噪声测量通常是用一个稍微不同的装置来实现的，这是为了使测量点噪声系数作为频率的函数成为可能，而不仅仅是测放大器在整个带宽上的总噪声系数。典型的设置是对图 14.6 所示的进行了一点修改，就是增加了一个混频器、本地振荡器和中频(IF)放大器，这就像超外差接收机里的结构[⑨]，如图 14.7 所示。

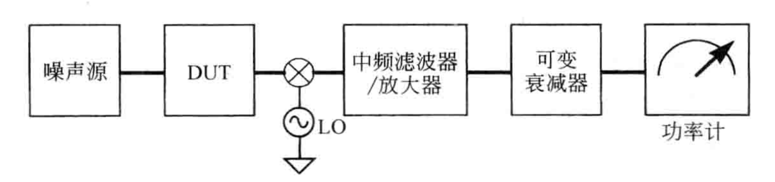


图 14.7　更为典型的 Y 因子测量设置

这里，通过 LO 频率的扫描来对 DUT 在不同频率下对噪声进行采样。那个 IF 放大器一滤波器的组合保证了这个噪声的测量是在某个窄带的、可控的，且以为 LO 设定的频率为中心的带宽中进行的。在许多实际实现中，DUT 和仪器之间会额外插入一个滤波器以限制带宽(比如，或许是为了衰减镜像响应)。

在以上任一方法中，如果将频谱分析仪用作功率计，那么就有必要在它前面放一个高增益、低噪声的前置放大器，因为频谱分析仪一般有很高的噪声系数(比如 30dB)，而这是为了保持良好的大信号线性度而做的设计折中。为了实现 NF 的大幅度降低，前置放大器的增

⑨　LO、混频器和中频放大器也都可以添加到图 14.5 所示的装置中，来改进信号发生方法，用这套系统很容易计算出点噪声系数。中频滤波器的噪声带宽决定了噪声点的频带宽度，而需要知道这个值以正确地计算点噪声系数。

益必须比分析仪的噪声系数大。总的噪声系数将接近于前置放大器自身的噪声系数。即使这样，前置放大-频谱分析仪这个组合的噪声系数依然是足够地高以至不能被忽略，而且级联噪声系数的 Friis 公式应该被用来修正测量得到的值。举一个具体数值例子，假设前置放大器的噪声系数和可用功率增益分别是 3dB 和 40dB，分析仪的噪声系数是 30dB。这个组合的噪声因数由 Friis 公式给出，

$$F = 2 + \frac{1000 - 1}{10^4} \approx 2.1 \tag{31}$$

或 3.2dB，这相比于 30dB 来说是一个很大的改进，其值现在只比前置放大器的固有噪声系数大一点点。最终当 DUT 连到这个组合上时，Friis 公式还要被用一次，3.2dB 现在被看作是第二级的噪声系数，而 DUT 的可用增益用在分母里。如果前置放大器的噪声系数是未知的，直接接热和冷的噪声源连到它的输入端，然后测量其噪声系数。一旦这个组合被采用这个方法进行了表征，它就可以用来确定 DUT 的噪声系数。

当使用频谱分析仪进行点噪声系数测量时，将分析仪的分辨率带宽设置为等于所需的点宽，并且尽可能减小检测到的信号的噪声。类似的，选择一个比分辨率带宽窄(好多)的视频带宽，以达到减小显示出的数据的噪声(回想一下，视频带宽控制着在检波器后面的输出信号的平均值)。这些考虑是从以下关系中得出的，这些关系适用于采用整流－平均型检波器的频谱分析仪：

$$\frac{\text{均方根输出噪声电压}}{\text{检测到的直流电压}} = \sqrt{\frac{b}{2B}} \tag{32}$$

其中 b 和 B 分别是视频和 IF 噪声带宽。

图 14.6 所示的基本设置也十分接近大部分现代自动噪声系数测试仪的核心部分所有的。比如，HP8970 包含了所有那些元件，并且增加了一个滤波器和前置放大器(就像频谱分析器的例子里)和在输入和输出端的一系列衰减器。有了这些额外组件，这个仪器能测量(并且校正)DUT 和噪声系数测量夹具的插入增益(损耗)。另外，为了完成准确的噪声系数测量，还必须知道计量计电路的噪声系数。下面是一个典型的使用商用仪器(特别是 8970B)的噪声系数测量的操作流程：

(1) 读出热/冷源的 ENR 校准值，将这些数据输入到仪器的内存中。8970B 的 ROM 中预存有一系列常用的校准频率，所以用户只需输入 ENR 值。

(2) 选择起始频率、结束频率和频率增量(步长)。

(3) 将噪声源连接到仪器上，以进行计量计的噪声系数测量，将频率设置到所要求的值，按下“校准”键开始校准过程。仪表依次激活和关闭噪声源来计算仪表的热和冷噪声功率：

$$P_{\mathrm{hm}} = k(T_{\mathrm{h}} + T_{\mathrm{em}})BG_{\mathrm{m}} \tag{33}$$

$$P_{\mathrm{cm}} = k(T_{\mathrm{c}} + T_{\mathrm{em}})BG_{\mathrm{m}} \tag{34}$$

这两个功率之比对于增益带宽积一点都不敏感，只有仪表的噪声温度是未知的：

$$\frac{P_{\mathrm{hm}}}{P_{\mathrm{cm}}} = \frac{T_{\mathrm{h}} + T_{\mathrm{em}}}{T_{\mathrm{c}} + T_{\mathrm{em}}} \Rightarrow T_{\mathrm{em}} = \frac{P_{\mathrm{cm}}T_{\mathrm{h}} - P_{\mathrm{hm}}T_{\mathrm{c}}}{P_{\mathrm{hm}} - P_{\mathrm{cm}}} \tag{35}$$

这个仪器在三个输入增益设置下测量出 T_{em} 的值。

8970 能够得到在每一个频率下进行的几次校准测量的平均值。测量的次数由“增加”

(increase)键控制。按住这个键,直到显示出期望的测量次数。这一步要先于“校准”模式的激活进行。

(4) 把 DUT 加在噪声源和仪器之间,选择“噪声系数和增益”。在实际能做到的范围,尽可能避免采用电缆。夹具越短越好——这是为了减小 DUT 之前的损失(以及因后续减法运算的不确定性而引入的任何误差),同时可以减小引入射频干扰(RFI)。这之后,仪器就可以测量级联(DUT+功率计)的热和冷功率如下:

$$P_{h,tot} = k(T_h + T_{e,tot})BG_m G_{DUT} \tag{36}$$

$$P_{c,tot} = k(T_c + T_{e,tot})BG_m G_{DUT} \tag{37}$$

这两个功率之比也对增益-带宽积不敏感,这里只有 DUT-功率计的组合的噪声温度是未知的:

$$\frac{P_{h,tot}}{P_{c,tot}} = \frac{T_h + T_{e,tot}}{T_c + T_{e,tot}} \Rightarrow T_{e,tot} = \frac{P_{c,tot}T_h - P_{h,tot}T_c}{P_{h,tot} - P_{c,tot}} \tag{38}$$

由这些噪声功率差之比,可计算出 DUT 的增益:

$$G_{DUT} = \frac{P_{h,tot} - P_{c,tot}}{P_{hm} - P_{cm}} \tag{39}$$

功率计的增益完全被消去了,所以它的值在理论上是不相关的。在计算出 DUT 的增益、功率计的噪声温度、功率计和 DUT 组合的噪声温度之后,就可以用 Friis 公式求解出 DUT 自身的噪声系数:

$$T_{e,tot} = T_{DUT} + \frac{T_{em}}{G_{DUT}} \Rightarrow T_{DUT} = T_{e,tot} - \frac{T_{em}}{G_{DUT}} \tag{40}$$

注意,这个计算结果只有当 G_{DUT} 等于可用增益时才是正确的。失配可能会使这两者不等,因而引入误差。

8970 这个测试仪也允许用户输入冷温度。默认的值是 296.5K,接近于典型的室温。

一个单独的夹具损耗(例如与一个网络分析仪一起的)的测量能修正任何 DUT 之前夹具的衰减。大部分仪器允许用户输入衰减值(比如通过 8970 的“衰减补偿”特性)然后仪器会自动减去这个衰减因子。这个仪器会把噪声温度转换成噪声系数,并且显示 NF 和 G_{DUT}。读这个描述的时间要比仪器实际进行测量的时间长多了。

14.7 误差源

有几种方式会使噪声系数测量出错。理解这些是什么是进行准确测量的关键。下面是一个短的列表,列出了常见问题、犯的错误和它们的解决方法。

14.7.1 外部噪声

经常会发生外部干扰耦合到测试装置。这些干扰可能是从电视、广播到数字设备(尤其是电脑及其显示器)中的 RF 源所发射的噪声。噪声系数测量最好在一个有屏蔽(“加金属丝网”)的房间里进行,以防止干扰入射到系统中去。如果没有这样的条件,一个无奈的第二选择是在干扰频率之外的下频段中进行点噪声测量,前提是干扰信号的带宽足够窄,使得这个方法可行。许多噪声系数测量系统提供了一个到示波器的连接来监视频谱。如果这样的

一个输出端不存在，可以采用一个普通的频谱分析器来取代。有了监视器的帮助，可以很容易地确定干扰引起的离散尖峰，然后可将测量频率适当地避开这些干扰。

14.7.2 夹具损耗

夹具带来的不规则的效应是没有尽头的误差来源。例如，一个恰当的噪声系数测量需要准确确定在 DUT 前面的存在的衰减（比如因电缆造成的衰减）。这个损耗（用 dB 表示）是要从测出的总 NF 中减掉的（当且仅当损耗是在 290K 时发生的）来得出 DUT 的真实 NF。如果损耗过大，那么最终结果的不确定性就相当大了，因为仪器将要用两个近乎相等的值做减法。例如，想象 DUT 前夹具的功率损耗是 20dB（一个相当大的值），并且 DUT 自己有 2dB 的噪声系数。噪声系数计将测出 22dB 的噪声系数，而这个值不可避免是有误差带的（比如说，0.5dB）。假设现在这个误差使测量的 NF 成了 21.5dB，对夹具损耗的单独测量可能也有相似的 0.5dB 不确定性，假设这种情况下我们测出 20.5dB。做了减法后，我们算出了 DUT 的 NF 是 1dB，而不是正确值 2dB，这是相当大的误差！实际上，对于噪声系数很低、DUT 前损耗又很大的放大器，很有可能会得到负值！因此，要谨慎对待采用数学方法去除大衰减的噪声系数测量。作为一般的规则，将任何这种 DUT 前损耗限制到比预期噪声系数小的值是可取的。这个损耗越小越好。

如果损耗不在 290K 测得的，那么这个带损耗的部件的噪声温度为

$$T_e = (L-1)T_L \tag{41}$$

其中，L 是衰减，T_L 是测量损耗的温度。这样得到的噪声温度可用来做精确的修正。

14.7.3 来自次级的贡献

另一个常见的错误是没有考虑到 DUT 后级电路的噪声（“次级贡献”）。一个有关联的考虑是所有的商业噪声系数仪都假设被测 DUT 增益是与可用增益一样的。如果 DUT 的输出阻抗非常大，与噪声系数仪的输入端不匹配，这个假设就是个一个不恰当的假设，而对次级贡献的计算就会是错误的，这可以从式(40)看出。

噪声源输出阻抗和 DUT 输入阻抗的不匹配情形也需要考虑。从 DUT 输入端返回到噪声源的反射，然后任何在噪声源出的不匹配又会在那里产生一个反射到 DUT。入射功率和这个被反射回来的功率的重叠可以使得从噪声源过来的功率与当匹配阻抗时是入射功率不同。使情况更为复杂化的是噪声源在冷和热模式可能具有不同的阻抗，这又加剧了误差，因为噪声系数和可用增益会因此改变[10]。

对所有这些误差的修正需要知道所有这三个不匹配，就如从下面的公式可以看出[11]：

$$K_G = \frac{(1-|\Gamma_s|^2)\,|1-\Gamma_1\Gamma_2|^2}{(1-|\Gamma_2|^2)\,|1-\Gamma_1\Gamma_s|^2} \tag{42}$$

式中，K_G 是测得的插入增益应该乘以的因子，以得到可用增益的正确值。这些反射系数分别指的是：Γ_1 是从噪声测量设备的输入端看进去的，Γ_2 是从 DUT 的输出端看进去的，Γ_s

[10] 参见 N. J. Kuhn. Curing a Subtle but Significant Cause of Noise Figure Error（去除噪声系数误差的一个微妙而有重要的来源）. *Microwave Journal Magazine*, June 1984.

[11] Fundamentals of RF and Microwave Noise Figure Measurements（射频与微波噪声系数测量基础）. Hewlett-Packard Application Note 57-61, July 1983.

是从噪声源的输出端看进去的。注意，如果所有这些反射系数都是零，K_G 即是 1。也要注意，为了进行恰当的修正，需要知道反射系数的幅度和相位。如果只有反射系数的幅度是已知的，我们能得到的最好结果是误差的上限。举一个后者的具体例子，假设 Γ_1，Γ_2 和 Γ_s 的幅度是 0.33，0.33 和 0.11，那么真正的可用增益可能是在 0.95～1.3 之间的一个数乘以所测得的插入增益。

14.7.4　噪声源校准的不确定性

噪声源的 ENR 的不确定性又是一个额外的误差源。如前面所述，噪声二极管必须对着标准进行校准。但校准从来都不是完美的，噪声二极管也不是完全稳定的（尽管市面上可得到的固态噪声二极管其稳定性是相当好的）。人们通常可以期待质量为仪器级的噪声二极管（如流行的 HP346B）在低频（如 10MHz）时具有 0.1dB 左右的 ENR 不确定性，而在更高的频率（18GHz）这个不确定性可能会增加到 0.2dB。随着 DUT 噪声系数的减小，这些不确定性表现出来的百分比误差显得更为严重了。

一个噪声二极管的输出在工作频率范围内不会完全是个常数，也不会遵循任何可追溯到第一性原理的简单的函数关系，因此噪声源的校准是在一系列离散频率值下进行的（频率个数的典型值是 10 或 20）。在校准点之间，你（或噪声系数测试仪）必须进行插值。当然实际的噪声输出可能和插值得到的值不尽相同，这又增加了另一个误差项。

14.7.5　冷温度不等于 T_0

还有另一个常见问题是冷噪声源温度很少会是 290K。二极管噪声源的冷温度等于周围环境温度，对大部分实验室来说，这比 T_0 要高 4～5℃。作为一个大致的经验准则，噪声源每比 T_0 高 1℃，测量得到噪声温度就要比真正的噪声温度要低 1℃。因为这个实验室温度较热的问题，通常测量是低估了 DUT 的噪声系数。对于更严格的修正，必须要准确测量这个冷温度值，并用式（10）来计算修正量。这个修正在噪声系数非常小时是十分重要的。考虑噪声二极管（也可能是其他源）在使用中变热也是很重要的，因此噪声源温度在测量过程中作为时间的函数会发生变化。

14.7.6　线性性的失效：二极管检波器

图 14.1 所示的直线提供了噪声系数的定义和测量方法的基础。如果被测器件是非线性的，噪声系数不能被唯一地定义。一个相关的例子是用作平方律检波器的二极管（由于历史原因，经常称作视频检测器）。在类似于这种的情形下，要用另外一个不同的品质因子来表征噪声性能。

正切信号灵敏度（TSS）就是这样一个指标。它的初始定义是对噪声的一个非常主观的评价：一个操作员在示波器上观察示波器在没有任何输入信号时的噪声输出，并记录下正向噪声尖峰的位置。然后打开信号，操作员调整信号幅度直到负向噪声尖峰的位置正好触到之前没有信号时的正向噪声尖峰。形式上讲，TSS 的定义就是产生这个条件的输入信号的电平。这个定义的问题是噪声是随机的，理论上有无界的尖峰。所以实验者要做相对随意的决定来判断何时这个尖峰相等发生，而且不同的实验者可能有不同的猜测（同一个实验者在不同的时间也可能做出不同的决定）。为了消除这个主观性，大部分二极管制造商将

TSS 定义为引起输出信号功率超过输出噪声功率 8dB 的可用输入信号功率。这些数字与输出功率比约 6.3 和电压比 2.5 相符。注意，如果输入功率在二极管表现为平方率器件的区域内时，那么一个比值为 2.5 的输出信噪电压比可由一个输入功率比为 2.5 而产生的。二极管的典型 TSS 值可能是－60dBm。

另一个品质因子是标称可测信号(nominal detectable signal，NDS)，定义为导致输出 SNR 为 1 的可用输入功率。TSS 和 NDS 一般都是频率和偏置电流的函数，所以为让 TSS 和 NDS 值有意义必须指明频率与偏置。

14.8 混频器的特殊考虑

当一个被测组件是(或包含)混频器时，存在的一个问题是应该采用单边带还是双边带的噪声系数测量。在绝大多数的情况下，单边带噪声系数是比较合适的选择，因为几乎没有通信系统会在主频段和镜像频段中都传输同样的信号。作者所知道的仅有的两个特例：一个是直接变频接收机，在那里主信号和它的镜像占用同样的频段；另一个是深空辐射探测系统，在那里噪声(即宇宙的噪声)本身就是信号。因为双边带噪声系数要小 3dB，专门玩弄指标把戏的事经常会发生，这个双边带指标当作单边带来发布。

为了更确凿地研究单边带和双边带问题，我们来考虑无线电工程师学会(Institute Of Radio Engineer，IRE，现在是 IEEE)的噪声系数定义，其分子上包含所有的输出噪声。而分母中仅包含与信号相关的噪声。如果只有一个边带含有信号，那么相关的频谱大致就会像图 14.8 所示。这里信号仅在带宽 B_2 范围内出现。从这张图可以看出，正确的噪声因子的定义为

$$F_{\mathrm{SSB}} \stackrel{\mathrm{def}}{=} \frac{N_{\mathrm{a}} + kT_0 G_1 B_1 + kT_0 G_2 B_2}{kT_0 G_2 B_2} \tag{43}$$

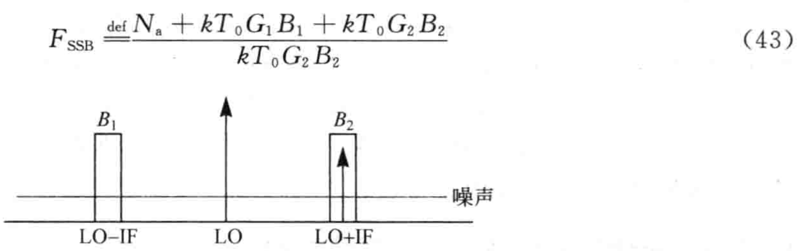


图 14.8　输入到混频器的单边带频谱

注意，这个公式允许接收机的两个带宽不相等，以及两个频带的转换增益不相等。

在罕见的双边带情况，期望的信号出现在两个边带内，如图 14.9 所示。此时相应的噪声因子为

$$F_{\mathrm{DSB}} \stackrel{\mathrm{def}}{=} \frac{N_{\mathrm{a}} + kT_0 G_1 B_1 + kT_0 G_2 B_2}{kT_0 G_1 B_1 + kT_0 G_2 B_2} \tag{44}$$

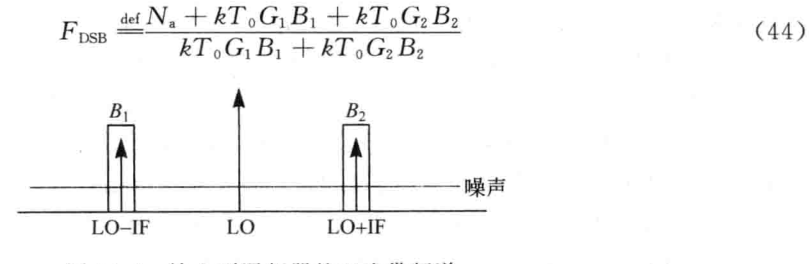


图 14.9　输入到混频器的双边带频谱

如果两个边带带宽一致并且转换增益也一样，那么正如前面所述，双边带噪声系数比单边带噪声系数值低 3dB。更一般的情况是，允许不同的转换增益(但仍假设带宽相等)，

$$F_{SSB} = F_{DSB}(1 + G_1/G_2) \tag{45}$$

在很多情况下，混频器前面加一个镜像压缩滤波器。在这个经常发生的情形中，把滤波器和混频器当作一个单元来处理是合适的。因为滤波器的作用就是在两个边带上产生一个不相同的变换增益，这样的话，双边带和单边带的噪声系数就没有－3dB 的差距了。

另一个重要的细微之处在于混频器端口的终断状况。因为一个混频器有 3 个端口：RF、IF 和 LO，任意端口的端断失配都会导致复杂的反射，从而致使测量结果不可信。一个很普遍的误差是接在一个无源混频器的 IF 端口端上的负载仅在差频处匹配，而在和频处却表现为高电抗的阻抗。尽管我们可能只关心差频分量，和频处的反射可能造成噪声和增益的病态行为。

此外，混频器的增益和噪声的特性通常随 LO 的功率变化而变化。为了使测量有意义，必须给定 LO 端口的功率。更好的方法是，在一个实际的 LO 频率范围内，将噪声系数(和转换增益)表达成 LO 功率的函数。最后，混频器通常对于 LO 端口的调幅噪声很敏感。它们将该噪声送到 IF 端口，形成另一个误差源。

14.9　参考文献

出自于惠普公司(Hewlett-Packard，现在是安捷伦，Agilent)的各种应用笔记是非常优秀的有关噪声测量信息来源。其中一些特别有兴趣的包括《精确与自动化了的噪声测量》(*Accurate and Automatic Noise Figure Measurements*)(HP 出版，64-3，1980 年 6 月)和《射频微波噪声系数测量原理》(*Fundamentals of RF and Microwave Noise Figure Measurements*)(HP 出版，57-1，1983 年 7 月)。另一个很好的信息源来自于 HP8970B 噪声系数仪的文档，其中详细地描述了该仪器的测量原理。

14.10　附录　两个粗略的目测方法

要准确测量噪声是相当花功夫的。通常，需要一个典型的噪声系数测试仪(或可能是一个频谱分析仪)来确定作为频率函数的噪声密度。对于相对较大量的噪声的快速评估，可采用一种粗略的测量方法。在这些情况下，仅需示波器和自己的眼球作为仪器。如果假设噪声满足高斯分布，那么峰峰值很少超过均方根值的 5～7 倍。因此，零阶的目测方法就是将有噪声的被测器件连接到示波器上，用目测来判断所显示的峰峰值大致是多少，然后除以 6 得到近似的均方根值。

当然，这个方法是非常粗略的，这在很大程度上是因为很难确定到底哪个是真正的峰峰值。由于示波器的亮度设置会影响到哪些地方显示为峰值，这就使情况更加复杂化。信号的轨迹越亮，则峰值显得越高。受睡眠、精神状态和咖啡因含量的影响，同样的操作者也会在不同的时间作出迥异的判定。

通过将测量转换成差分形式，一种目测技术的巧妙扩展方法能消除大部分的不确定性[12]。这里，噪声信号同时驱动双踪示波器的两个通道，在采用交流扫描模式而不是交替模式(chop mode，这是为了避免在两个扫描间通过示波器的断续振荡器引入的关联性)操作。以足够大的初始位置的差距开始，在两条轨迹之间会存在一个暗带。调节位置控制钮直到暗带刚消失，这时两条轨迹合成一条模糊的带，并且此带的亮度从中心向外逐渐变暗。注意这个描述暗示着结果独立于绝对的亮度。去掉噪声信号，然后测量两条基线之间的距离。得到的值近似为电压均方根的 2 倍。采用这种方法可以取得 1dB 的绝对精度。

这种方法可行的原因是因为当两个全同的高斯分布的和当两个分布分开一个恰好为均方根值 2 倍时，会出现一个最大平坦化的顶。

因为眼睛对对比度的判断不是很完美，所以暗带消失时不可能用极高的精度作出判断。当采用上述的步骤时，大部分人在暗带真正开始消失前就感觉到它好像已经消失了。这个不确定性引起的误差对于大多数人来说都在 1dB 量级。因此，如果你非常挑剔，那么可能要将测量结果减去 0.5dB。另一种选择是采用两种不同的方法测量噪声，一种是用以上给出的步骤，另一种是让开始时两条轨迹彼此重叠。对于后者的这个初始条件，调节间距直到较暗的区域刚开始出现。取两个读数的平均，并计算出两个读数之差作为不确定性的度量。如果测量时多加小心再加上一点练习，小于 1dB 精度的可重复性是不难取得的。

---

⑫ G. Franklin and T. Hatley. "Don't Eyeball Noise"(不需要目测噪声). *Electronic Design*, v. 24, 22 November 1973, pp. 184-187.

第15章 振荡器

15.1 引言

考虑到人们需要花这么大的力气去消除在大多数反馈系统中的不稳定性，因而看起来构建一个振荡器会是十分简单的事。然而，根据墨菲定律(译者注：墨菲定律说的是，事情如果有变坏的可能，不管这种可能性有多小，它总会反着你的愿望而发生)，事情没有这么简单。这正如你为了让老天下雨而去带一把伞一样。射频工程师中有一个笑话：所有的放大器都会发生振荡，而所有的振荡器都会仅是个放大器。

在本章中，将讨论振荡器设计中的几个方面。首先，说明为什么纯线性的振荡器是无法在现实中实现的。然后，给出了一种线性化方法，这种方法利用描述函数来提供对非线性效应是如何影响振荡器性能的深入理解。有了描述函数，预测振荡器的频率和幅度会变得很直截了当。

本章中还包括了对谐振器技术的回顾，并且将重新讨论锁相环电路，不过这一次是从频率合成器电路的背景下来进行分析。在本章结束时，我们对振荡器的结构作一个综述。相位噪声这一重要问题则在第 17 章中详细分析。

15.2 与纯线性振荡器关联的问题

在负反馈系统中，我们通过寻求大的正相位裕度来避免电路的不稳定性。那么对于制作一个振荡器，似乎只要取得零或负相位裕度就可以了。可以利用图 15.1 中所示的正反馈系统根轨迹图来较为仔细地考察一下这个思路[①]。

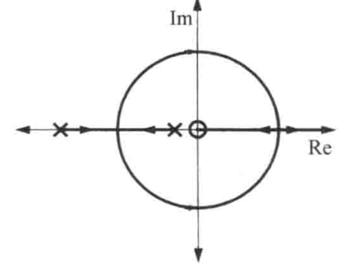


图 15.1 振荡器例子中的根轨迹图

这个轨迹曲线在振荡器的设计过程中反复出现，因为它适用于一个具有两个极点的带反馈的带通谐振回路。如轨迹图所示，对于环路增益幅度的某一特定值闭合回路的极点正好就在虚轴上。其相应的冲击响应因此就是一个既不随时间衰减也不增长的正弦信号，这样看起来我们就有了一个振荡器。

但是，上面这个说法存在着几个实际问题。首先，振荡信号的幅度是依赖于冲击信号幅

---

① 对于那些不太熟悉根轨迹的读者：根轨迹只是在某些参数发生变化时系统的闭环极点所可能遍历的所有可能的值。通常来讲这个变化的参数是环路增益，但它也可以是其他参数，例如环路传输极点频率。有关根轨迹的细节可以参考 T. H. Lee. The Design of CMOS Radio-Frequency Integrated Circuits. 2nd ed.. Cambridge University Press, 2004。

度的(别忘了,这是一个线性系统)。这一特性通常说来并不是我们所希望的,因为几乎是在所有的情况下,我们希望振荡器产生的信号幅度是一个常数,不随着初始状态的改变而改变。另一个问题在于如果闭合回路的极点不是精确地位于虚轴上,那么振荡信号的幅度要么会不断衰减,要么就会不断增大。

这些问题对任何采用纯线性方法来进行振荡器设计都是不可避免的。因此解决这些问题的办法就是有意识地利用非线性效应,**现实中的所有振荡器电路都是依赖于非线性原理的**。要理解为什么非线性效应能够解决这些问题,并且培养在振荡器电路中直觉分析和设计能力,下面引入描述函数这一概念。

15.3 描述函数

我们已经看到系统的线性化描述通常是足够的,即使是对于一个非线性系统来说。例如,双极型晶体管的增量(即交流小信号)模型就是来自于器件的指数性转移特性的线性化。只要激励是"充分的小",系统具有线性响应行为这一近似就可以很好地被满足。

另一种将输入输出转移特性线性化的方法就是在频域范围内进行线性化。具体来说,考虑对一个非线性系统施加一个具有特定频率和特定幅度的正弦输入信号的情形。输出信号通常是由很多不同频率和不同幅度的正弦信号叠加而成。系统的线性表述可以通过去掉除了与输入信号频率相同以外的所有其他输出信号分量来得到。一个包含了所有可能的由输入信号到保留下来的输出信号的相位变化及幅度变化率的集合就构成了对这个系统的非线性性的描述函数。如果输出信号的频谱主要由基频分量构成,则由描述函数所得到的结果将是一个具有相当精确的近似。

为了进一步确保下边将要进行的一系列分析的正确性,我们还要对非线性系统加上以下这些限制:它们不产生输入信号的分频谐波(即频率比输入信号频率低的谐波,直流信号就是一种分频谐波)。之所以要加这个限制的理由很快就会变得清楚。对于射频系统而言,这一限制不像它初看起来那么严格,因为我们常常可以用带通滤波器来消除分频谐波与高次谐波分量。

作为一个产生描述函数的具体例子,我们考虑一个理想的比较器,它的输入输出满足下边的关系:

$$V_{\text{out}} = B \operatorname{sgn} V_{\text{in}} \tag{1}$$

如果用一个频率为 ω 幅度为 E 的正弦波来驱动这个比较器,那么输出信号就会是一个方波信号,频率与输入信号相同,幅度为一常数 B,与输入信号的幅度无关。而且输出信号与输入信号同时跨过零值点(因此没有相位偏移)。这样,输出信号可以用下边的傅里叶级数来表示:

$$V_{\text{out}} = \frac{4B}{\pi}\sum_{1}^{\infty}\frac{\sin\omega nt}{n}, \quad n\text{ 为奇数} \tag{2}$$

只保留基频分量($n=1$),并且将输出信号与输入信号的幅度相比就得到了这个比较器的描述函数:

$$G_{\text{D}}(E) = \frac{4B}{\pi E} \tag{3}$$

因为在这个特别的例子中描述函数的相位没有发生偏移，函数也与频率无关，它仅仅依赖于输入信号的幅度。

需要注意的是，这个比较器的描述函数表现出增益与驱动信号的幅度成反比的关系，这与线性系统中增益与驱动信号的幅度无关的特性是截然不同的。我们很快就会发现，在一个系统中，这种反比例增益特性在提供负反馈以稳定振荡的幅度上是极其有用的。

15.3.1　描述函数的一些例子

在已经说明了如何得到一个描述函数的步骤之后，我们在图 15.2～图 15.4 中不加推导地给出一些常用的非线性系统的描述函数[②]。在图 15.4 中，R 的值必须要小于 1，否则施密特触发器不会被触发，其输出将为直流，要么为 B，要么为 $-B$。

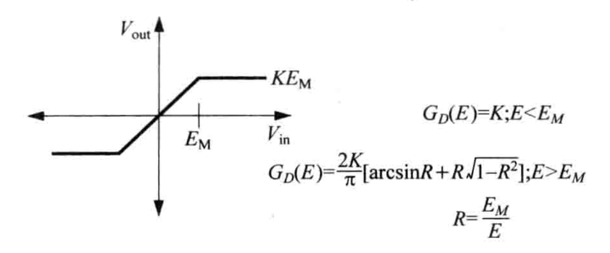


图 15.2　饱和放大器的转移特性与描述函数

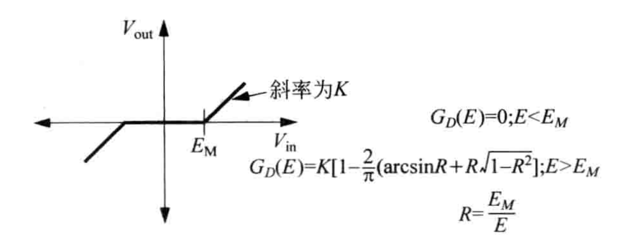


图 15.3　带跨越失真的放大器的描述函数

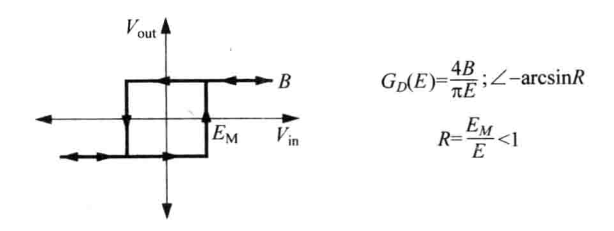


图 15.4　施密特触发器的转移特性和 G_D

值得注意的是，描述函数本身都是线性的，即使它们所描述的系统是非线性的（明白了吗）。因此，叠加原理成立，描述一组非线性性和的描述函数就等于对各个非线性性描述函数的相加。这个特点对于我们得到其他的非线性系统的描述函数是非常有用的，而不仅仅

② 例如，参见 J. K. Roberge 的佳作，*Operational Amplifiers*（运算放大器）一书，Wiley，New York，1975

针对上边所列出的几个例子。

15.3.2 对于晶体管与真空管的一个通用描述函数

尽管前节所给出的描述函数是十分有用的，但是与射频电路设计更相关的是表述由一到两个晶体管组成的电路的描述函数，因为表征射频电路工作的甚高频率信号是很难由一个由许多晶体管构成的电路来产生的。

为了说明一个普遍的分析方法，参看图 15.5 中的电路。假设电容的值足够大，以保证在频率 ω 下可以看作是交流短路，并且晶体管是一个理想的晶体管。我们会在调谐振荡器中用到这个电路。由于谐振回路的带通滤波作用，这使得我们的描述函数分析可以导致精确的结果。

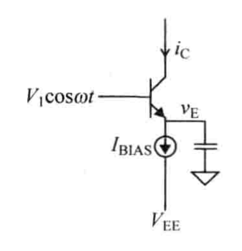


图 15.5　大信号跨导器

在开始大信号跨导（也就是描述函数）的详细推导之前，先来定性地预言一下结果的大致情况。随着幅度 V_1 的增加，发射极电压 V_E 也被拉到了一个更高的值。当输入信号达到最大值时，V_E 也大致达到最大值。在基极电压从峰值开始下降不久，晶体管就处于截止状态了，因为输入信号的下降速度超过了电容通过电流源放电的速度。因为电流源在每个周期内对电容进行放电，基极-发射极结在输入信号回到峰值之前又被正向偏置，这引起了一个集电极电流的脉冲。这个过程不断重复，所以集电极电流由一系列的脉冲电流组成。

值得注意的是，为了定量地得到在大驱动幅度信号下的大信号跨导我们不需要知道再多的集电极电流的具体形状。唯一有关联的事实是在这个大信号极端状况下，集电极电流是由一系列窄峰构成的，就像是图 15.6 中所示的假想的基极、发射极电压和集电极电流一样[③]。

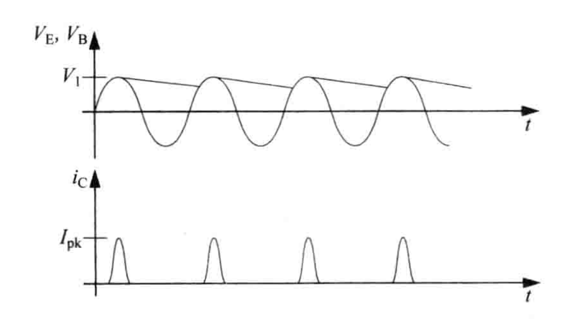


图 15.6　在大输入电压下的假想的发射极与基极电压及集电极电流

无论电流的波形是什么样子，基尔霍夫电流定律要求电流的平均值应该等于 I_{BIAS}。也就是，

$$\langle i_C \rangle = \frac{1}{T}\int_0^T i_C(t)\,\mathrm{d}t = I_{BIAS} \tag{4}$$

③　在这里“假想的”这个词是“不正确”这个词的委婉说法。但是尽管所示的波形并不是严格正确的，所得到的结果和推论是严格正确的。特别是，这张图帮助我们理解为什么在大的驱动幅度下双极晶体管与 MOSFETs（包括长沟与短沟）的描述函数本质上是与 JFETs 乃至真空管相一致的。

现在,集电极电流的基波分量的幅度由下式给出:

$$I_1 = \frac{2}{T}\int_0^T i_C(t)\cos\omega t\,dt \tag{5}$$

尽管不知道 $i_C(t)$的具体表达形式,但是我们知道它是由一系列在大信号驱动极限下的窄脉冲组成的。而且这些电流脉冲大致出现在输入为最大的时候,所以余弦函数在冲击电流存在的一小段时间内可以近似地被当作单位量(即 1)。于是

$$I_1 = \frac{2}{T}\int_0^T i_C(t)\cos\omega t\,dt \approx \frac{2}{T}\int_0^T i_C(t)\,dt = 2I_{BIAS} \tag{6}$$

这说明在 V_1很大的极限下,基波分量的幅度近似地等于偏置电流的两倍。于是描述函数的幅度为

$$G_m = \frac{I_1}{V_1} \approx \frac{2I_{BIAS}}{V_1} \tag{7}$$

需要强调的是,前面的推导过程中都没有用到过晶体管特性的具体细节。因为推导过程并不是针对特定的器件假设,所以式(7)可以通用在 MOSFET(长沟和短沟)、双极晶体管、JFETs、GaAs MESFETs、甚至是真空管中。

在推导式(7)的过程中,我们假设了驱动幅度 V_1的值是“大”的。为了定量地给出这个概念,对于长沟和短沟的 MOSFET(比如说 VMOS RF 功率器件)以及双极晶体管,我们来计算 G_m/g_m这个比率。

对于长沟器件,g_m与 I_{BIAS}的比率可以写成

$$\frac{g_m}{I_{BIAS}} = \frac{2}{V_{GS} - V_t} \tag{8}$$

于是

$$\frac{G_m}{g_m} = \frac{V_{GS} - V_t}{V_1} \tag{9}$$

很显然,对长沟 MOSFET 来说,V_1“大”的定义是相对于$(V_{GS}-V_t)$而言的。对于短沟器件我们来重复这个过程,得到[④]

$$\frac{g_m}{I_{BIAS}} = \frac{2}{V_{GS} - V_t} - \frac{1}{E_{sat}L + (V_{GS} - V_t)} \tag{10}$$

这个公式在很短沟道的极限下,给出的值正好是长沟器件值的一半。因此,

$$\frac{V_{gs} - V}{V_1} \leqslant \frac{G_m}{g_m} < \frac{2(V_{gs} - V_t)}{V_1} \tag{11}$$

最后,对于双极型晶体管,有公式

$$\frac{g_m}{I_{BIAS}} = \frac{1}{V_T} \tag{12}$$

从而

$$\frac{G_m}{g_m} = \frac{2V_T}{V_1} \tag{13}$$

对于双极型晶体管来说, V_1“大”是相对于热电压 kT/q 来定义的。

---

④　在这里,我们用到了在注解 1 引用的 Lee 这本书的第 5 章介绍过的短沟 MOSFET 的近似解析模型。

尽管G_m公式的推导仅在V_1大的情况下成立，但是实际的振荡器通常是能满足这个条件的，所以这个限制并不如想象的那么严格。在下一章中我们还会看到大的V_1也是减小相位噪声所非常需要的，所以可以这样说，一个设计得好的振荡器是会自动地满足这个近似所需要的条件的。虽然如此，仍然应该注意G_m永远不会超过g_m，所以要注意不要不正确地应用诸如式(13)的公式。为了强调这一点，图 15.7 近似地描绘了G_m/g_m的实际曲线相比于从式(13)所得到的曲线。尽管这个等式只是严格地用于双极型晶体管，但是图 15.7 所示的情况对于其他器件也是成立的。

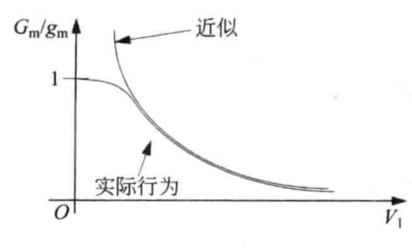


图 15.7 G_m/g_m与V_1的关系曲线

上边已经列举了不少描述函数的例子，下面我们来举一个例子来说明如何应用描述函数来分析振荡器。

15.3.3 实例：COLPITTS(考尔匹茨)振荡器

类似于多形式波形发生器(function generator)这样的弛张振荡器很少用在高性能发射机中，因为它们产生的信号频率纯度不够高。调谐振荡器则应用得更加广泛些，其主要原因只有等到我们学完相位噪声以后才能明白。现在，暂把调谐振荡器的优越性当作一个公理来接受。我们目前的注意力先集中在应用描述函数来预测一个典型的调谐振荡器的输出幅度，例如图 15.8 中所示的 Colpitts 电路[⑤]。在本章中，我们将会看到很多仅仅在一些细微的地方有所不同的振荡器是用它们的发明者的名字来命名的。遵循惯例，我们保留了这些振荡器的命名方式，但是建议读者把注意力集中在工作原理上而不是名称上。

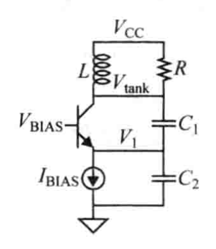


图 15.8 Colpitts 振荡器(偏置电路未画出)

设计这些振荡器的基本方法是十分简单的：把一个谐振电路与有源器件相连接。Colpitts 振荡器最明显的特点就是用了一个用电容分压的谐振回路，再通过有源器件提供的正反馈来实现振荡。在图 15.8 中，电阻 R 代表了总的负载，不管是由于谐振腔的有限 Q 值、晶体管的输出电阻，或者其他由振荡器所驱动的负载(我们假定振荡器的输出总会被用在某个地方)。在实际的使用中电流源常常被换成一个普通的电阻，这里用电流源仅仅是为了稍微简化分析。

从我们推导描述函数的过程来看，晶体管的特性可以当作一个大信号的跨导G_m来看。为了简单起见，我们忽略晶体管的所有动态元件(译者注：即电容)，以及所有的寄生电阻，而在一个精确的分析中这些其他的元件都是应该考虑进去的。当然，这个晶体管也有一个大信号的源栅电阻，这在模型中必须加以考虑。从描述函数中可以得到启发，我们可以把这个电阻定义成源电流中的基频分量与源栅电压的比值。我们实际上已经知道了这个比值，它就是$1/G_m$。因此可以把振荡器电路化成图 15.9 所示的模型。

⑤ Edwin Henry Colpitts 在 1915 年初构想了这个振荡器，此时他在 Western Electric 工作。他的同事 Ralph Vinton Lyon Hartley 在一个月之前(即 2 月 10 日)演示了他自己的振荡器(后被称为 Hartley 振荡器，译者注)。

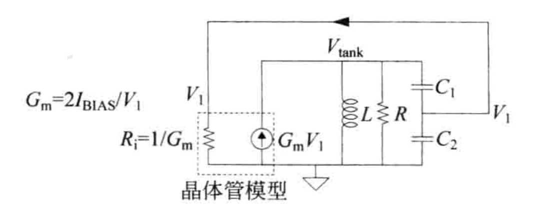


图 15.9　Colpitts 振荡器的描述函数模型

为了简化接下来的分析，首先把电容分压器当作一个理想的变压器，将输入电阻 R_i 折算到主谐振回路两端，这样得到了一个带有正反馈环的简单 RLC 谐振回路。注意，得到的这个简化电路与谐振回路在谐振频率下没有相位差，因而对这个特定的例子，谐振腔的谐振频率也就是振荡器的工作频率。同时还要注意到在任何时候，受控电流源都产生一个正弦的输出，其幅度为 $G_mV_1=2I_{BIAS}$。所以可以用具有相同幅度的一个独立正弦电流源来代替。通过这些观察，我们把电路变成了如图 15.10 所示的形式，图中 C_{eq} 是两个电容的串连值

图 15.10　Colpitts 振荡器的简化模型

$$C_{eq}=\frac{C_1C_2}{C_1+C_2} \tag{14}$$

以及

$$\omega=\frac{1}{\sqrt{LC_{eq}}} \tag{15}$$

同样地，R_{eq} 也是谐振回路中原有的电阻与反映过来的晶体管输入电阻的并联[⑥]：

$$R_{eq}\approx R\,||\,\frac{1}{n^2G_m} \tag{16}$$

式中 n 是电容分压器上的电压分配系数

$$n\overset{\text{def}}{=}\frac{C_1}{C_1+C_2} \tag{17}$$

幅度 V_1 就是将谐振回路的幅度 V_{tank} 乘以电容分压器的电压分配系数，于是可写出

$$V_{tank}\approx\frac{V_1}{n} \tag{18}$$

现在我们已经准备好了往下进行推导所需要的所有公式。在谐振的状态下，谐振回路上电压的幅度就是电流源幅度与谐振回路净电阻的乘积：

$$\begin{aligned}V_{tank}&\approx\frac{V_1}{n}\approx 2I_{BIAS}R_{eq}\\&\approx(2I_{BIAS})\left[R\parallel\frac{1}{n^2G_m}\right]=(2I_{BIAS})\cdot\frac{R}{n^2G_mR+1}\end{aligned} \tag{19}$$

化简最后得到

⑥　在这个以及其他有关的等式中，用“约等于号”是因为我们将电容分压器当作了理想的阻抗变换器。在只有当这个分压器所在的电路的 Q 值很高时，这样的近似是合理的。

$$V_{\text{tank}} \approx 2I_{\text{BIAS}}R(1-n) \tag{20}$$

于是，振荡器的幅度也就直接正比于偏置电流和谐振回路的等效电阻。由于晶体管输入电阻所引入的谐振回路负载的效果通过系数$(1-n)$得到考虑，这样，可以通过调节两个电容的比值来控制负载效应。因为R同时还控制着Q，通常说来会被设置成尽可能大，而调节I_{BIAS}也就成为控制幅度的主要手段。

作为一个特定的数值例子，假设图15.11中所描述的是一个60MHz的振荡器电路。考虑电路中所给定的元件参数，电容分压比n大约是0.155[⑦]。预期的振荡幅度大约是1.4V。实际测量了一个双极晶体管的电路得到幅度约为1.3V，与理论的预期符合得很好。需要强调的是，这个结果与构成这个振荡器的有源器件的类型在很大程度上无关。这个预期结果原来是为MOSFET设计所作的，但是与用双极型器件实现的实验结果符合得相当好。

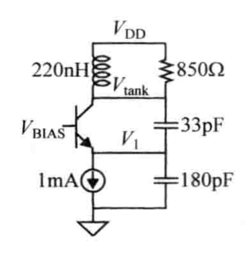


图15.11 Colpitts振荡器例子

15.3.4 COLPITTS设计细节

我们现在来考虑一个1GHz的采用微带线实现的Colpitts振荡器的实现来补充上面那个低频的例子。同时，要设计一个压控振荡器，而不是一个只能输出单一频率的振荡器。

有许多的振荡器的拓扑结构在理论上几乎是无法区分的，然而经过汇聚的进化，仅有不多的几种选择被广泛使用。一个考虑是，并不是所有振荡器都能方便地加偏置或者能提供好的调谐范围。例如，可以选择将晶体管的集电极接地或者将其连到某个正的V_{CC}的拓扑结构。也可以选择将基极接地的拓扑结构。这当中的每一个选择都会导致不同程度的在实现偏置上的难度以及对于寄生效应的敏感性的问题。

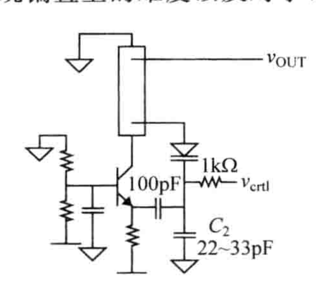


图15.12 略加简化了的VCO示意图

图15.12中的Colpitts结构在许多情形下都工作得不错(译者注，图中的v_{OUT}输出端应该连接到集电极，而非地)。这里我们选择将集电极直流接地，并且用负电源电压来对晶体管进行偏置。这样一个极性的选择来自几方面的考虑：将谐振腔直接接地可以避免需要将集电极的电源交流旁路(采用扼流圈或者BFC)，同时它不需要输出耦合电容。这些简化使得很容易将电感的一端用良好的短路线连接(例如，将电感放置在PCB板的边缘处，这样可以充分利用这个结构的优势，而不需要采用通孔)。另外，因为许多PLL产生的控制电压是用地端作参考电位的，因此可变电容可以连接到用接地点为参考的负载上，进而避免了需要加偏置电路的问题。

在这个结构中，一个适当长度的微带线充当电感的角色。此电感与可变电容和C_2以及发射极电流并联的电容进行串联得到的电容一起实现谐振。从而，调谐范围取决于可变电容与其他电容的比例。幸运的是，在这个具体的结构中，可以使得可变电容占主要的成分。

⑦ 在实际应用中，一般当C_2/C_1的比率大约为4时电路的相位噪声性能最好，这对应于$n=0.2$。这个大致的规则可以用第17章讨论的时变理论给出一个更严格的解释。

这可以通过选择一个足够大的 C_2 来实现。这也就是我们为了最大化调谐范围所想要的值。与其他一些 Colpitts 结构不同，C_2 两端的寄生电容影响相对很小。然而，实际中我们无法任意加大 C_2，因为它也是反馈电压分压的一部分。如果它过大，则反馈将减弱，从而振荡会消失。将 C_2 设为与可变电容最大值接近是一个可行的初始选择。可变电容的最大值进一步由施加在变容管两端的电压的最小结电压决定。例如，如果 VCO 由 PLL 控制，而 PLL 的控制电压在 0.5～4.5V 之间，则 0.5V 时的结电容就是可变电容的最大值。

另外一个实际的考虑是将基极旁路路径做的尽量短。与基极串联的任何电感都会使电路振荡在一个你想不到会发生的频率下。甚至在这些寄生的振荡模式产生之前，它们也会以不希望有的方式对振荡器产生干扰。

负载电感设计

假设一个在调谐频率范围中心时的净电容为 8pF（假设可变电容最小值为 8pF，最大电容为 18pF，这个可变电容再串联上 C_2 的 22pF，再加上 4pF 的发射极电容及其他的寄生电容，然后采用这两个总电容值的几何平均值计算得到[⑧]），此时你需要一个约为 3nH 的电感。通过手工的反复试的手段来得到这个值的电感有点儿困难，但不是不可能。需要特别小心地使可变电容连接过程中引入的寄生电感。如果成功地实现了这个电感，则计算得到的调谐范围大约是标称（中心）频率的正负 15%。

如果有需要既减少电感设计遇到的问题，同时又增大调谐范围，可以最大化可变电容两端的控制电压，从而减小可变电容的最小值。然而，这通常意味着需要设计与构造一个放大器来对 PLL 芯片输出的控制电压进行适当的放大。

偏置

从描述函数分析中我们知道，振荡幅度是偏置电流和损耗（包含了负载电阻的作用）的一个函数。其中损耗可以在一定程度通过改变抽头位置的方式由设计者控制。要保持负载电感的长度尽可能短的要求限制了抽头改变的灵活性，所以选择一个合适的偏置电流在设计中就变得相当重要。取决于抽头的位置，偏置电流在几个毫安范围内的变化会得到输出为 0～7dBm。在设计有关偏置电路版图的过程中，在有可能会影响谐振频率或者调谐范围的地方，需要尽量减小寄生电容的影响。就跟其他带有抽头的谐振糟路一样，负载结构本身并不能保证对输出阻抗的实部和虚部同时实现控制，因此需要额外的匹配网络。

另一个考虑是虽然可变电容的阳极是接到其值为零的直流电压，但也会存在着一个交流成分。因此，控制电压需要比阳极电压的峰值要高得足够多，从而避免可变电容会处于正向偏置。如果这个问题发生了，则振荡可能会停止，进而造成反馈回路处于混乱状态。因此对谐振糟路的幅度存在着一个限制。

这个简化的电路图没有显示出偏置电路的元件值。在设计时，开始可以尝试令射级电阻压降为 1～2V，然后选择基极偏置分压器使得其电流约为集电极电流的 1/10。这个粗略的经验规则在多数情况下可以确保可被接受的偏置稳定性，但这需要在各种关心的情况下加以确认验证。由于 β 变化引起的偏置稳定性问题可以通过增加流过基极偏置分压器的电流得到改善，而 V_{BE} 变化引起的稳定性问题可以通过增加射级电阻压降得到改善。

---

⑧ 这是一个对于总的“其他”电容的一个悲观估计。实际情况会有差别，尤其是在采用晶体管取代可变电容的情况下。

其他需要注意的事项

谐振槽路的电感可能是导致设计困难的一个来源，这是因为它的长度短。你可能会推测，线宽越窄则调谐范围越好，因为这样对于一个给定的阻抗可以采用一个更短的导线作为电感，这又使得这个导线更加接近一个纯电感而不是一个准谐振器。毕竟线宽对于谐振槽路损耗的影响只是个二阶的效应，因为大部分 FR4 PCB 板的损耗来自于介质，所以在这个场合采用窄的导线可能不是个问题。

然而，大部分是由于实际实现上的原因，采用短的线会产生一些问题。为了允许用长一点的导线，可考虑采用比通常使用的宽度更宽的导线。但需确保导线的所有方向的尺寸要远小于 $\lambda/4$，否则强的分布效应会使你的计算不正确（或者会迫使你采用更复杂的计算公式）。在实际设计中，最好选择一个长度，它不会导致电感在谐振频率的某些整数倍频率处出现高阻，因为这些失真乘积项不会被谐振槽路滤掉（记住，你所做的“电感”实际上不过是就一个终端短路的传输线）。然而，在这个实验中，并没有对失真的指标，所以你不需要过多考虑这个问题。我们在这里提到这个是希望你在投身到实践时，不会抱怨那些在象牙塔里的老师从来就没有提到这一点。

另外一个要动点脑筋的问题是测到的调谐范围太窄，而变容管的容差又太大，结果是 VCO 的中心频率不一定是你想要的值。在这种情况下，如果这个频率范围距所需的频率足够近，不一定需要抛弃现有电感重新来过。如果是要略微提高频率，可以在电感靠近地的一端加上一些铜从而降低电感值。如果是要略微降低频率，可以将电感导线削窄一些。为了进一步降低频率，可以将一小段金属薄片焊在导线上用于增加电容（位置越靠近集电极，效果越明显）。

如果输出功率太低，可以将偏置电流增加。如果输出功耗在调谐范围内变化太大，可以使电感上的输出端抽头更靠近集电极的部位抽。电感的损耗，也就是其等效电阻，随着频率改变。如果这个与频率相关的损耗占主导地位，则输出幅度会随着频率显著变化。将输出端在靠近集电极的部位抽出来，则谐振器的负载越重，但这时的负载更加接近一个常数。增加偏置电流可以补偿平均的输出功耗的跌落。需要在环路增益、输出功耗平坦性和滤波质量（更重的负载意味着变差的 Q）三者之间进行折中。

起振，二阶效应，以及某些细致分析

在上边的分析中，没有特别提到确保振荡器起振的条件。但从图 15.1 所示的一般化的根轨迹图来看，很明显一个必要条件是在小信号下的环路传输系数大于 1。为了计算在启动的时候是否会遇到问题，应该将跨导设置为晶体管处于小信号状态下的值（正振荡器起振前，电路确实处于小信号工作状态），并以此计算环路传输增益。如果这个增益没有超过 1，那么振荡器将不会启动。要解决这个问题，需要综合调整偏置电流、器件尺寸以及分压比率。

对于刚刚分析过的例子，让我们来确定一下为保证振荡启动的最小跨导值。这个最小跨导 g_m 以及偏置电流一起决定了器件的宽度。我们用图 15.13 所示的模型。当谐振的时候谐振回路两端的电压幅度为

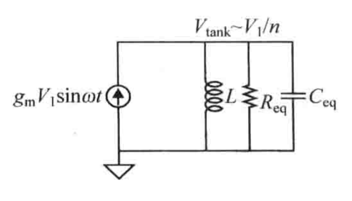


图 15.13　Colpitts 振荡器的启振模型

$$V_{tank} = \frac{V_1}{n} = g_m V_1 R_{eq} = g_m V_1 \left[R \parallel \frac{1}{n^2 g_m} \right] \qquad (21)$$

由此计算出所要求的最小跨导的公式为

$$g_m > \frac{1}{R(n-n^2)} \tag{22}$$

当 $n=1.55$，$R=850\Omega$ 时，得到绝对最小的跨导大约为 9mS。但是应该注意，仅仅满足未振荡时单位值的环路增益的跨导并不足以构建一个好的振荡器。此外，描述函数仅仅在信号幅度大的情况下才准确，因此只有当小信号跨导要远大于大信号情况下的跨导值时，电路才会正确工作。对于第一轮初步设计过程而言，选择 g_m 为最小值的 5 倍是比较合理的。因此，我们将把小信号跨导设计成 45mS。

我们来估算需要的器件宽度，开始时假设栅的过驱动电压足够小以使得器件的特性满足平方律的关系。这样可以利用式(8)来计算栅过驱动电压：

$$\frac{g_m}{I_{BIAS}} = \frac{2}{V_{gs}-V_t} \Rightarrow V_{gs}-V_t \approx 44\text{mV} \tag{23}$$

这个驱动电压与 $E_{sat}L$（比如说，1～2V）这个典型值相比较确实相当小，因此我们继续假设下边的计算是基于长沟器件的。计算这个假设下的 W/L 值，在使用典型的迁移率和 C_{ox} 的情况下，得到的值大约为 6000。对于 0.5mm 沟长的器件来说，宽度大约为 3000mm，这是一个相当大的值。这个沟宽是在比较低的偏置电流的情况下得到的，当采用更高的偏置电流时，一定可以使用沟宽小得多的器件。

在前面的推导中除了忽略了起振条件以外，还用了其他一些简化的假设来减少推导过程中的繁杂度。例如，晶体管的寄生参数都被忽略了。现在我们来考虑如何加入这些因素来修正我们的分析。

栅源（原文有误，这里已订正，译者注）（基极-发射极）电容和漏-体（集电极-地）电容都是并联在谐振回路上的，要修正它们的一阶效应只要简单地减少外接电容保证振荡器的频率不变就可以了。但是这些电容是非线性的，所以如果它们构成了谐振回路电容的相当大的一部分，失真是不能令人满意的。谐振槽路的温度漂移特性也常常会因此受到影响。

源栅电容（发射极-基极）和源极与衬底（发射极-地）之间的电容是直接并联在 C_2 的两端的，类似的情形也发生在其他的器件电容上。

另一个必须要考虑的因素是晶体管的输出电阻，因为它也构成对谐振回路的负载。许多高速晶体管的 Early 电压很低（比如，10～20V 或者更低），所以这个负载影响有时候是十分严重的。在一些糟糕的情况下，必须使用共基-共发放大器（或者某些等效的方案）来解决这个问题。在其他的情况下，只需要在计算时计入这个负载的影响就可以了，以便更加精确地预计振荡幅度。

关于幅度问题最后一个要提及的点是要强调因为幅度的反馈控制在振荡器的工作中是一个基本的过程，所以幅度的不稳定性总是存在的。也就是说，幅度不再是一个稳定的值，而是以某种方式发生变动（比如准正弦变化）。这种情况称为“间歇性振荡”，是振荡器设计中的令人头痛的事。为了对间歇性振荡有更好的了解以及提出防止间歇性振荡的方法，我们采用了与分析其他反馈系统的稳定性一样的分析方法。也就是说，我们采用环路传输、跨越频率，以及相位裕度等概念。主要的微妙之处在于需要用射频信号的包络来描述这些量。另一个微妙之处在于幅度控制的非线性导致线性分析仅仅在工作点附近有效。基于这些考虑，可以在方便的地方将环路切断（仍要保持所有的负载，正如在分析其他环路传输中必须做的一样），然后在切断的位置施加一个射频信号。这个测试信号的幅度需要与环路实际工

作时该点的幅度一致，从而保证我们在与实际工作时一致的情况下进行稳定性的评估。由于幅度控制的非线性，需要谨慎地考察不同幅度时的环路传输，从而可以确定（或者阻止）间歇性振荡发生的幅度范围。

下面，可以考察振荡器在时域或者频域的表现。对于前一种，考察幅度发生阶跃变化时环路传输的响应。为了考察频域包络环路传输，在切断点注入一个经正弦调制的射频载波，并且扫描调制的频率，然后观察输入调制信号导致的输出调制信号的增益和相位的变化。

对于谐振槽路接在漏端的调谐振荡器，一个自然的选择是在漏端切断环路。在这一点上，向谐振器注入一个射频电流，并且使该射频电流幅度发生一个阶跃变化。谐振器自身会对阶跃变化提供一个等效的单极点滤波，而到晶体管源端的电容耦合会有进一步的动态影响。

为了考察这个过程的细节，先来考虑一个 Colpitts 滤波器。为了简化分析，先构造一个等效电路，如图 15.14 所示。这两个电路的等效性可以通过比较环路传输来确认。漏端的连接是一个很方便的切断点来进行这个环路传输比较。我们看到只要电路图中的元件反映了所有的器件寄生参数，则两个电路是完全等价的。

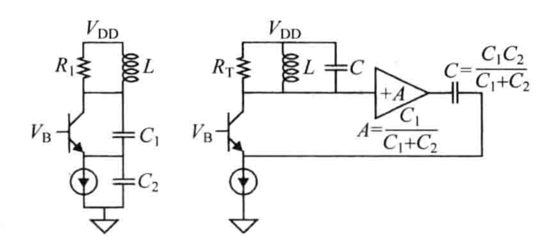


图 15.14　Colpitts 振荡器及计算包络环路传输系数的等效电路模型

有了计算包络环路传输的等效电路，可以将环路传输分解成独立的成分。首先，分析图 15.15[⑨]中的容性耦合电路。这里，负载电流包括直流和射频成分：[⑩]

$$i_L(t) = I_{DC} + i_{dc}(t) + [I_O + i_o(t)]\cos\omega t \tag{24}$$

图 15.15　带有非线性负载的电容耦合电路

直流和射频电流幅度中不变的部分分别为 I_{DC} 和 I_O，对应着射频驱动幅度 V_1。将驱动幅度进行一个扰动 $v_{in}(t)$ 一般而言会产生三个效应。第一个效应是会使电容两端电压的直流成分变化 $v_c(t)$，这是因为非线性负载的整流效应。如果射频输入电压的幅度发生变化，则直流的电容电压通常也会发生变化。

这个整流效应同时也会改变通过非线性负载的直流电流，其变化用 $i_{dc}(t)$ 表示。最终，

⑨　此分析在 Kenneth K. Clarke 和 Donald T. Hess 的 *Communication Circuits: Analysis and Design*（通信电路：分析与设计），Krieger，Malabar，FL，1994，得以讨论。

⑩　这里"直流"的用法不是严格的，是用于将包络与射频两个成分相区分。

流入非线性负载的射频电流的幅度会发生一个变化 $i_o(t)$。这个幅度的变化是直接由 $v_{in}(t)$ 以及流过负载的直流电流的变化导致的。

我们希望确定小信号导纳 $i_o(s)/v_{in}(s)$，然而要得到这个导纳并不是很容易的事。注意，对这个问题的描述本质上是假定了小信号分析是成立的。如果只考虑 $|v_{in}(i) \ll V_1|$ 的情形，那么可以将 $i_{dc}(t)$ 和 $i_o(t)$ 每一个表示为 $v_c(t)$ 和 $v_{in}(t)$ 的一个简单线性组合。经过 Laplace 变换，可以得到

$$i_{dc}(s) = G_{00}v_c(s) + G_{01}v_{in}(s) \tag{25}$$

$$i_{out}(s) = G_{10}v_c(s) + G_{11}v_{in}(s) \tag{26}$$

其中各种常数 G_{mn} 是稍后要确定的电导。

因为依然有以下等式成立：

$$i_{dc}(s) = -sCv_c(s) \tag{27}$$

可以将 $i_{dc}(s)$ 的两个表达式等同起来以得到

$$-sCv_c(s) = G_{00}v_c(s) + G_{01}v_{in}(s) \Rightarrow v_c(s) = \frac{-G_{01}}{sC + G_{00}}v_{in}(s) \tag{28}$$

这个表达式给出小信号 DC 电容上的电压即是简单的小信号 RF 输入幅度经低通滤波器滤波的版本。

将式(28)代入式(26)就得到所寻求的在输入包络电压与输出包络电流之间的小信号关系：

$$i_o(s) = G_{10}\frac{-G_{01}}{sC + G_{00}}v_{in}(s) + G_{11}v_{in}(s) \Rightarrow \frac{i_o(s)}{v_{in}(s)} = G_{11} - \frac{G_{01}G_{10}}{sC + G_{00}} \tag{29}$$

经过一些重新安排，上式变为

$$\frac{i_o(s)}{v_{in}(s)} = \frac{G_{11}(sC + G_{00}) - G_{01}G_{10}}{sC + G_{00}} = G_{11}\frac{\left(\dfrac{sC}{G_{00}} + 1\right) - \dfrac{G_{01}G_{10}}{G_{00}G_{11}}}{\dfrac{sC}{G_{00}} + 1} \tag{30}$$

到目前为止我们所做的推导都是完全一般性的，例如式(30)不限于 MOSFET 或者双极性晶体管。不需要知道不同电导的具体值，我们只说明涉及的导纳由一个极点和一个零点组成。这个结果从物理上是讲合理的，因为有一个储能元件(从而对应于一个极点)。另外，电容提供了一个对高频成分(载波和包络都在内)的串通馈路，这是一个零点所表现的行为。这样，这个到目前为止给出的分析能够通过微观上的合理性测试，因此是令人满意的。

现在我们转向确定这些电导都是些什么的任务，注意这些比例常数的定义都是由式(25)和式(26)的频域的版本中得到的，

$$G_{00} \stackrel{\text{def}}{=} \left.\frac{di_{dc}}{dv_c}\right|_{v_{in}=0} \tag{31}$$

$$G_{01} \stackrel{\text{def}}{=} \left.\frac{di_{dc}}{dv_{in}}\right|_{v_c=0} \tag{32}$$

$$G_{10} \stackrel{\text{def}}{=} \left.\frac{di_o}{dv_c}\right|_{v_{in}=0} \tag{33}$$

$$G_{11} \stackrel{\text{def}}{=} \left.\frac{di_o}{dv_{in}}\right|_{v_c=0} \tag{34}$$

注意，这几个电导中至少有两个是我们熟悉的。从定义中可以看出 G_{00} 只是非线性负载的小信号直流电流和直流电压的比值，从而它也就是一般意义上在偏置点处的小信号电导。同理，G_{11} 是射频输出电流的幅度除上固定的电容电压上的射频输入电压变化的幅度。从而 G_{11} 就是非线性负载的描述函数电导。

两个包括一个直流项和一个射频项的比例的电导之前并没有遇到过。其中一个是 G_{01}，也就是电容直流电压不变的情况下，整流后的直流电流除以电容两端射频输入电压幅度的变化。另一项 G_{10} 是射频输入电压幅度不变情况下，射频输出电流幅度的变化除以电容直流电压的变化。

我们需要做的最后一件事是对于与漏端相连的谐振槽路的包络行为的一个量化的描述。具体地说，考虑一个正弦驱动电流幅度的阶跃变化。谐振槽路上电压的包络响应会表现为对于阶跃电压的单极点低通滤波响应。由于一个 RC 低通滤波器的单边带宽为 $1/RC$，可以预期对应于 RLC 带通滤波器的时间常数为 $1/2RC$[①]。在谐振时，漏端负载提供的包络阻抗由下式给出：

$$\frac{v_{\text{tank}}(s)}{i_{\text{o}}(s)}=\frac{R_{\text{T}}}{2sR_{\text{T}}C+1} \tag{35}$$

因此完整的环路传输系数为

$$A\frac{i_{\text{o}}(s)v_{\text{tank}}(s)}{v_{\text{in}}(s)i_{\text{o}}(s)}=A\left[G_{11}\frac{\left(\dfrac{sC}{G_{00}}+1\right)-\dfrac{G_{01}G_{10}}{G_{00}G_{11}}}{\dfrac{sC}{G_{00}}+1}\right]\frac{R_{\text{T}}}{2sR_{\text{T}}C+1} \tag{36}$$

注意，这里有两个极点和一个零点。同时也要注意到一个不好的现象，即包络环路传输的符号是正的。虽然正反馈本身不一定会导致不稳定，仍必须避免环路传输的幅度过大。

对应于这个环路传输一个可能的根轨迹解释了为什么间歇性振荡会发生，如图 15.16 所示。因为我们并不确切知道零点的位置，另一个可能性是如图 15.17 所示。在这个例子中，极点都不是复数，但其中一个极点可能会取正的实部。如果任何极点在右半平面，包络会变得不稳定。如果在右半平面有复数极点对，则不稳定性会表现为一个(准)正弦调制。如果极点为实数，则这个调制会具有松弛的特征。

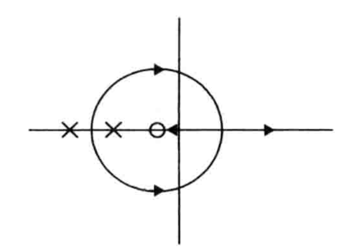


图 15.16　包络反馈环路的可能的根轨迹图

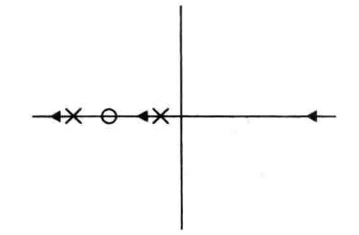


图 15.17　包络反馈环路的另一个可能的根轨迹图

① 强调单边带宽的目的是更好地说明低通和带通滤波器的相似性。一般习惯用从 DC 到正的 −3dB 频率来描述低通滤波器的带宽，而不是用正的 −3dB 到负的 −3dB 作为低通滤波器的带宽。对应于单边带的时间常数控制了上升时间，从而也就是 RC。对于带通滤波器，其单边带宽是 $1/(2RC)$，意味着控制包络上升时间的极点时间常数为 $2RC$。

可惜的是,如果没有给定一个具体的非线性负载,无法得到定量的结论。更可惜的是,对于实际的 MOSFET,得到一个准确的推导几乎是不可能的(对于双极型晶体管已经难度够大了)。从而,多数的对于间歇性振荡的评估都需要在分析中进行仿真。然而,在仿真中得到 G_{01} 和 G_{10} 是很容易的[12]。这样,即使对这些参数的简单的解析表达式并不是很容易得到,采用仿真的方法却可以不是很困难地得到它们的实际值。

即使不进行这样的仿真,如果间歇性振荡发生了,也可以确认一些通用的策略来防止间歇振荡发生。借鉴一般的放大器设计中的主极点补偿的思想,可以考虑增加谐振器的 Q 值。Q 值增加导致的带宽降低意味着它对幅度控制环路贡献的极点移到了一个更低的频率(即变得更加占主导)。这迫使跨越频率更低,而这意味着更大的相位裕度。

如果谐振器的带宽无法降低,仍可以用其他方法降低幅度控制环路的跨越频率。可以通过改变电容抽头比的方法减小反馈信号从而减小环路传输。当然也可以结合几种不同的方法来达到目的。

一个可能存在的困难是这些参数中很多是相互依赖的。例如,取决于如何实现谐振槽路 Q 值的增加,谐振槽路的包络阻抗会升高,从而会导致环路传输系数幅度的升高,这会导致电路趋向不稳定。从而对于特定的电路,需要更细致的考虑来选择一个最好的策略。

在十分难处理的情形下,可能需要加外部的幅度控制(例如,直接测量幅度,将其与一个参考值比较,从而适当地调节偏置电流,如图 15.18 所示)。这种将幅度控制和基本的振荡分离的方法不仅为解决稳定性问题引入了一个新的自由度,而且使得振荡器设计时不需要为诸如启动的可靠性(或速度)及幅度稳定性这些因子间作出折中。

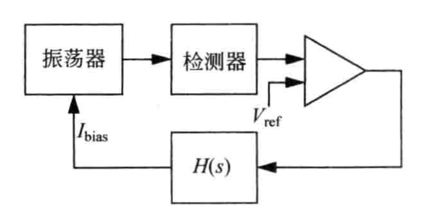


图 15.18 带有独立幅度控制的振荡器

还有两个描述不希望的振荡器行为的项是频率牵引(frequency pull)和电源推进(supply push)。这两个概念都是描述由于寄生参数引起的振荡频率的漂移。牵引效应可能来自很多的因素,从负载变化到与其他周期信号的寄生耦合。采用缓冲及其他的隔离技术可以减小牵引效应。

电源推进反映了振荡频率与供电电压并非完全无关,这是人们不希望发生的现象。例如,器件电容可能会随着偏置电压变化,导致振荡频率的漂移。电源推进可以通过选择谐振槽路元件的数值大到可以将寄生元件的影响压倒,以及采用稳压电源等手段来进行抑制。对后者进行很好的滤波是十分重要的,尤其是需要注意电源电压中的 $1/f$ 波动,因为这会导致振荡器的(close-in,近场)相位调制。

15.4 谐振器

前面讨论的描述函数的例子分析了调谐振荡器。因为调谐电路天然具有带通滤波的作用,失真和噪声相对于基本的频率成分都得到了衰减,所以毫不奇怪,这些电路的性能是与

[12] 读者可能有理由来问为什么不直接仿真整个振荡器。答案是间歇振荡频率可能会远低于振荡频率,从而仿真时间会很长。采用更加严格的分析可以确定应该进行怎样的仿真。

共振器的质量密切相关的。在我们对振荡器电路进行细致的讨论之前，首先介绍一些共振器技术。

谐振器技术

λ/4 谐振器

除了常见而颇受青睐的 *RLC* 谐振电路以外，还有很多种方法来制作谐振器。随着频率的提高，对于谐振器来说要得到足够高的 *Q* 值也变得越来越难，这是因为要求所用到的元件值常常变得不现实。

一种可用的替代方法是采用由末端短路的 λ/4 传输线做成的谐振器。对于偏离谐振频率一个小的量的频率，这个结构的行为十分像一个平行的 *RLC* 网络。所要求的物理尺寸对于 UHF 频带或者更高的频率来说用分立形式来实现是可行的。举一个例子，300MHz 频率的波在真空中的波长约为 1m，所以 λ/4 共振器的尺寸大概是 10in(25cm)。在 FR4 电路板上，这差不多是 6in(15cm)长。

当频率为几个 GHz 的时候，尺寸将可以与用集成电路实现相适应。例如，在 3GHz 时，真空中的波长大约是 1in(2.5cm)。用现在通常用到的电介质层，λ/4 集成电路谐振器大约是 $\frac{1}{2}$in 或更小。

但是短路的传输线与集总元件的 *RLC* 谐振器还有很重要的一点不同：对于基波的奇次谐波来说传输线的阻抗都表现为无穷大(至少是非常大)。有时候这个特点是我们所期望的，但是这也有可能导致振荡器同时输出多个频率或者无规则地从一个频率跳变到另一个频率去。因此也许需要一些附加的谐振组件来抑制我们不需要的振荡模式。

石英晶体

最常用的非 *RLC* 谐振器是用石英制作的。石英晶体用在无线电收发装置上的优良特性和潜力是大约在 1920 年由贝尔实验室的 Walter G. Cady 首次发现的[⑬]。石英晶体是一种压电材料，可以在机械应变和电荷之间进行转换。当电压被加在一条石英片的两侧面的时候，晶体就会发生形变。当一个应变被施加在它上面的时候，石英晶体的表面又会出现电荷[⑭]。

被用在无线电频率[⑮]下的大多数石英晶具有体切变振动模式，见图 15.19。在这个模式下，共振频率与石英板的厚度成反比关系，这是由图中的简略公式[⑯](用的是国际单位制)给出的。

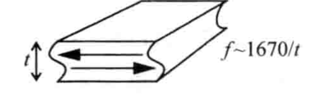


图 15.19　晶体的体切变模示意图

尽管石英并不具有特别大的压电效应，但是对于射频电路来说它有其他一些特别有价值的特点。最为突出的就是材料出

⑬ W. G. Cady. The Piiezo-Electric Resonator(压电谐振器). *Proc IRE*, v. 10, April 1922, pp. 83-114. 他的第一个振荡器比较复杂，是基于一个二端口的压电滤波器。

⑭ 压电的机-电转换由 Jacques 和 Jacques Curie 与 Pierre Curie 两兄弟发现的(Pierre Curie 后来与 Marie Sklodowska 相识并结婚，Marie Sklodowska 也就是居里夫人)。参考"Developpement, par pression, de l'electricite polaire dans les cristaux hemiedres a faces inclines" [Development, by Pressure, of Electrical Polarization in Hemihedral Crystals with Inclined Faces](在切有斜面的半面形晶体上，通过压力来产生电极化), Comptes Rendus des Seances de l'Academie des Sciences, v. 92, 1880, pp. 294-295. 他们的朋友，物理学家李普曼(Gabriel Lippman)后来基于热动力学预测了这个效应的反效应的存在，并且很快就被 Curie 兄弟证实。

⑮ 用在电子手表中的晶体用到了一种扭转的振动模式使得较小的尺寸就可以在较低的频率下振荡(32.768kHz)。

⑯ 这个公式忽略了其他维度的影响。

色的稳定性(包括电学的和力学的特性)。而且,在特别的角度[17]上切割石英晶体可以得到非常低的温度系数。此外,石英晶体几乎可以无损的转换能量,Q 值的范围从 $10^4 \sim 10^6$。[18]

图 15.20 给出了一个石英谐振器的电学模型。电容 C_0 代表与接触面及引线相关的并联电容,而 C_m 和 L_m 代表了力学的能量存储。电阻 R_s 表征了在任何真实系统中都会存在的损耗。

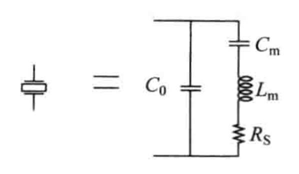


图 15.20 晶体的符号与模型

制作良好的晶体的电阻在一个非常粗略的近似下是跟振荡的频率成反比的,通常有如下公式:

$$R_S \approx \frac{5 \times 10^8}{f_0} \tag{37}$$

这是一个半经验的公式,只有在没有测量结果的时候才应该用这个公式[19]。

如果给出了 R_S,Q 以及共振的频率,C_m 和 L_m 的值可以计算出来。一般来说,因为石英晶振的 Q 值非常的高,我们计算出来的等效电感值也非常大,而串联的电容的值就几乎接近于零值。例如,一个 Q 值为 10^5、工作频率为 1MHz 的晶振的等效电感值约为 8H,C_m 的值约为 3.2fF。在这里石英晶振显示出了相对于集总元件式 LC 实现的极大的优越性,在任何情况下,集总元件是不可能有这样极端的值的。

大约在高于 20~30MHz 的情况下,石英板厚度的要求变得容易实现了。例如,一个工作在基频为 100MHz 的石英晶振大约只能有 17μm 厚。尽管如此,如果能利用更高的振动模式的话,晶体的厚度还是可以制备的。边界条件的要求使得决定了只有奇次的倍频信号是可能存在的。因为一些其他的原因,倍频信号并不严格的是基频信号的整数倍(但是很接近,一般高出不到 0.1%)。三倍频和五倍频的晶振还是相当常见的,七倍频甚至九倍频的振荡器偶尔也能见到。但是,随着倍频倍数的提高,想要让振荡器工作在我们想要的模式下也变得越来越困难了。

另一个非常粗略的规律是,有效的串连电阻是按照倍频倍数的平方关系增长的。因此

$$R_S \approx \frac{5 \times 10^8}{f_0} N^2 \tag{38}$$

式中 f_0 是指 N 阶倍频后的工作频率。

因为倍频信号的频率并不严格是基频信号的整数倍,所以晶体也必须在要求的倍频信号处对准确的频率来进行切割。切割得好的倍频晶体能够提供与基频晶体相当的 Q 值。

石英晶体的制造工艺也是一门极其先进的艺术,谐振频率的误差被控制在 50ppm[20] 以内的晶体不难被制作。如果支付更高的费用,还可以得到更好的结果。惰性的化学性质使得石英晶体的性质相当稳定。通过合适的选择,配合上主动或者被动形式的温度补偿或者控制,还可以得到温度系数好于 1ppm/℃ 的石英晶振。因为这些原因,石英晶振几乎普遍

---

[17] 它也可以控制得到非零温度系数的晶体。这种技术已经被用来制作温度频率变换器,并用在普通电路达到不了的极端温度。

[18] 在较低的频率上,受空气阻尼的作用,Q 值大大降低。那个高端的 Q 值是相应于放置在真空中的晶体。

[19] 这个公式只能严格地用在"AT-cut 切割"的晶体的基频工作状态下(译者注:At-cut 是一种最常用的石英晶体的切割方法,其产生的石英条的朝向使得其工作在厚度方向切变振动模式,有很好的操作特性)。

[20] $1\text{ppm}=10^{-6}$。

地存在于所有的通信设备和装置中(更不必提低廉的腕表了，其误差仅有 20×10^{-6}，这样的误差能保证连续工作一个月也只会差 1min)。

表面声学波(SAW)器件

因为石英晶体是工作在体振动模式下，所以高工作频率要求非常薄的石英板。例如，一个工作在 1GHz 的基频模式下的石英晶体只有 1.7μm 厚。除了显而易见的制作上的困难，薄石英板在加了过大的电激励之后也十分容易破碎。因为它们有很高的 Q 值，故只需要加一个不大的电驱动，就可以很容易地得到一个大的振荡幅度。即使在实际破碎之前，因为极度的形变也会引起许多我们不希望出现的非线性效应。

摆脱这些限制的一个办法就是利用表面声学波，而不是体声学波。如果材料可以提供表面声学波模式，那么有效的厚度就会大大减小，可以通过实际可加工得到的晶体尺度实现更高的共振频率。

铌酸锂($LiNbO_3$)就是一种能提供表面声学波而损耗很小的压电材料，现在已经广泛被用来制作一些工作在石英晶体所达不到的频率下的谐振器和滤波器了。遗憾的是，能得到石英晶振工作频率精确控制的廉价方法不再有了。但是这个性能依然足够满足一些大批量、低成本的应用，例如自动控制车库门。它的典型工作频率典型的是 250～300MHz，以及移动电话的前端滤波器。

令人失望的是，无论是石英晶体还是铌酸锂都与传统的 IC 制造工艺不兼容。遗憾的是，硅是没有任何压电效应的，这也很。用在 IC 制造中通常采用的材料层，制作不出本征是高 Q 值的谐振器。

15.5 调谐振荡器分类

正如马上会变得清楚的，用一到两个晶体管及一个谐振器组合构成振荡器的方法看起来是没有限制的。在下边的这些例子中，仅仅提供最主要的一些方案，读者也许会问哪一种技术更好。实际上，通常说来对于一个给定的应用来说，这里的任何一种设计方法都可以得到足够好的结果。然而当考虑相位噪声问题时，恰当的选择就会有更多的限制了。

15.5.1 基本 *LC* 反馈振荡器

这些振荡器的基本成分很简单：一个晶体管加上一个共振器。许多振荡器都是以第一个发明该拓扑结构的作者的名字来命名的。然而，正如我们将要看到的，可以采用一个一定程度上统一的方法来描述所有的这些振荡器。

我们已经遇到过 Colpitts 振荡器的其中一个版本。Colpitts 振荡器的另一种形式是反馈信号是从发射极返回到基极的，而不是从集电极到发射极。就是说，晶体管可以被接成射极跟随器或者普通的共发射极放大器。而这两种方式都有一个纯正反馈回路。

Hartley 振荡器(如图 15.21 所示)本质上与 Colpitts 振荡器是相同的，但是采用了电感抽头分压的形式而不是电容分压。Hartley 振荡器很早就出现在无线电装置中了，因为中间抽头的电感随处可以得到。但它现在不常见了。

图 15.21 Hartley 振荡器(未画出偏置电路)

原则上，也应该能用有中间接头的电阻分压器，但是这种专门的结构并没有一个特别的名字与之相联系。

Clapp 振荡器（如图 15.22 所示）是 Colpitts 振荡器的一种变化形式，将单个电感用电感与电容的串连来代替[21]。Clapp 振荡器实际上就是一种在电容分压器链上增加了一个抽头的 Colpitts 振荡器，就像是在图 15.23 中所等效的样子。这个额外的抽头允许电感上的电压（以及分压器链上的）大大超过在集电极或发射极上的电压，因此也就能够得到超过电源电压，甚至是超过器件的击穿电压。更大的信号能量能够帮助我们克服许多噪声过程（特别是相位噪声，这将会在第 17 章中进行了讨论）的影响，以便改进信号的纯度。

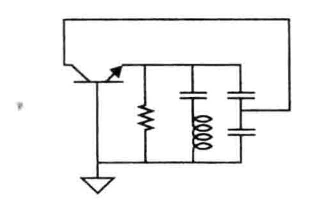


图 15.22　Clapp 振荡器（仍然未标出偏置电路）

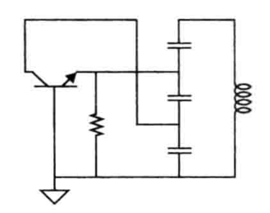


图 15.23　重画过的 Clapp 振荡器

比起其他拓扑结构，Colpitts 振荡器是最常遇到的。它使用的抽头电容是与 IC 生产的工艺最相兼容的，但是电感就不是如此。Colpitts 振荡器得以广泛流行的另一个原因是因为它极佳的相位噪声特性，我们很快就会看到这一点。

另一种振荡器的存在实际上是依仗于某种调谐式放大器的不稳定性。回想一下，一个共发放大器工作在比谐振频率低的频率上时（因此谐振槽路呈现出电感性），它的输入阻抗可能会是负的[22]。负的电阻可以用来补偿另一个谐振回路上的损耗，从而产生振荡。TITO 振荡器（图 15.24）用到了一个密勒耦合电容。在很多设计中（特别是在非常高的频率下），并不需要专门加一个耦合电容，器件本身所固有的反馈电容就可以满足产生所需要的负阻。这个发现突出了在高频情况下，对非单向放大器设计中同时在输入和输出上使用调谐电路的难度。

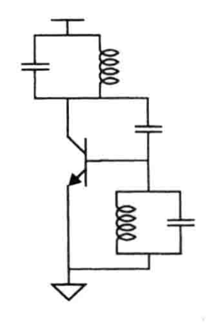


图 15.24　调谐输入-调谐输出（TITO）振荡器（不完全的偏置电路）

因为有两个调谐电路，理论上 TITO 振荡器可以产生高频谱纯度的信号。但是它需要两个电感，这在 IC 制造工艺上是不希望的。另一个不利于这种电路被采用的因素是需要仔细调谐两个谐振回路以将其置于恰当的工作状态。

[21] 见 James K. Clapp. An Inductive-Capacitive Oscillator of Unusual Frequency Stability（一种具有不寻常的频率稳定性的电感性—电容性振荡器）. Proc. IRE, v. 36, 1948, pp. 356-8. Clapp 发明这个电路的时候在 General Ratio 公司工作。

[22] 但是单纯满足这个条件并不够。

15.5.2 晶体振荡器的集锦

可以辨认出许多晶体振荡器都是从 LC 振荡器演变过来的。比如，如图 15.25 所示，处于串联谐振模式（这时晶体呈现出低阻态）的晶体被用到谐振回路的反馈馈路上，以使得让整个电路只在我们想要的频率上才形成闭合的反馈回路。

由于晶体上的寄生电容（C_0），并联在晶体两端的电感通常（但并不是总是）在实际应用中是需要的，以防止因为这个晶体的并联电容所提供的反馈而造成在我们不想要的频率下的振荡。这个电感与电容谐振，使得仅仅由晶体的串联 RLC 支路来控制反馈。

Colpitts 晶体振荡器的另一个版本画在图 15.26 中。在这个特定的结构中，谐振回路的电容分压器提供了如同经典的 Colpitts 振荡器反馈路径。但是，只有对于晶体的串联谐振频率上的信号来说，晶体才将基极短接到了地上，这样就使得反馈回路中仅仅在那个串联频率下有足够的增益来维持振荡。当晶体的一端必须接地的时候，这种拓扑结构是相当有用的。

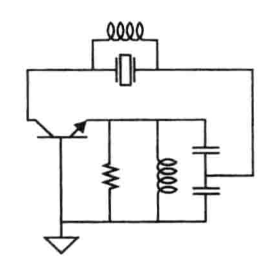


图 15.25　Colpitts 晶体振荡器

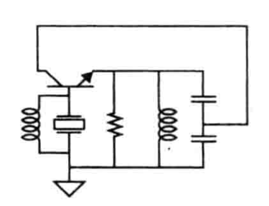


图 15.26　经过改动的 Colpitts 晶体振荡器

还有一种布局结构是 Pierce 振荡器[23]，表示在图 15.27 中。在这种振荡器中，假设图中的电容已包含了晶体管与寄生的电容，这样晶体管本身就是一个理想器件。给定这个假设，要严格满足零相位差的要求就需要振荡器的振荡频率要比晶体的串联谐振频率稍高一些。这就是说，在振荡器的工作频率上，晶体是表现出电感的特性来。这个性质的好处就在于不需要另外再接一个电感到电路中，就可以让振荡器工作（射频电路中的扼流圈可以用一个大电阻或者电流源来代替）。因此，它比 Colpitts 振荡器更容易实现集成，尤其是在一些低频应用中。

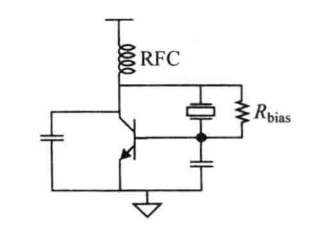


图 15.27　Pierce 晶体振荡器

---

[23] 无线电的开拓者，企业家，哈佛大学教授 George Washington Pierce 在无线领域作出了很多贡献，以他命名的晶体振荡器只是其中一个例子。参见 G. W. Pierce. “Piezoelectric Crystal Resonators and Crystal Oscillators Applied to the Precision Calibration of Wavemeters”（用于波长计的压电晶体谐振器与晶体振荡器）. *Proc. Amer,. Acad. of Arts and Sci.*, V. 59, October 1923, pp. 81-106. 以及美国专利＃2133642，1924 年 2 月 25 日申请，1938 年 10 月 18 日获批。他在 Cady 为他演示了一个早期的压电振荡器之后很快作出这些发明。除了他关于振荡器的工作之外，他将点接触二极管命名为晶体整流器并且对一个早前的基于热学的解释进行批驳。他在 1909 年写的教科书 Principles of Wireless Telegraphy（无线电报基础）是有关这个技术出现的早期最畅销的书。

晶体为什么必须呈现感性这一点可以理解如下。如果要实现零相位差，而且晶体管的跨导已经引入了 180°的相位偏移，那么剩下的无源器件必须提供另外 180°的相位偏移。有两个极点的 RC 网络无法提供 180°相位偏移(相当接近，但是不是等于，这依然无法保证工作)，所以晶体必须起到电感的作用。

因为 Pierce 晶体振荡器的输出信号的频率并不等于晶体共振的频率，所以必须这样来切割晶体，以使得当与一个特定的负载电容(这种情况下，两个电容是串联的)一起时振荡器的频率等于需要的频率。

作为一个对 Pierce 振荡器最后一点说明是，它恰好形成了许多"数字"振荡器的基础。例如，当被设置工作在线性区时(比如接一个很大的反馈电阻，就如在一个单一晶体管的实现中那样)，一个普通的 CMOS 反相器可以作为一个带有增益的电路组件工作。只要加入适当的输入输出电容，将晶体连接在输入输出之间，就有很大的可能得到了一个振荡器。通常来说，要得到一个全部的 CMOS 摆幅，这个电路还需要有一到两级的缓冲(这当然就需要更多的反相器)，同时也起到了隔离振荡器的核心部分与负载的作用。

15.5.3　其他的振荡器结构

在一些应用中，还需要同时输出两个正交的输出信号。一个可以自然提供(至少在原则上)两路正交信号的振荡器在反馈回路中用到一对积分器(见图 15.28)。根据满足幅度的条件，可以推导振荡器的工作频率为

$$\omega_{osc} = K \tag{39}$$

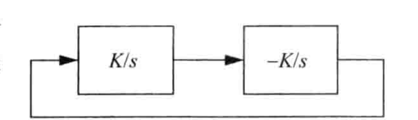


图 15.28　正交振荡器方框图

这样，调谐也许会被积分器的增益所影响。此外，所需要的正交关系可以从任何一个积分器的两端得到。

在实践中，未得到考虑与建模的动力学效应导致与理想情况的偏离。举个例子来说明这一点，考虑附加极点对根轨迹的影响。不再是由一对纯虚数构成的情况，有了附加极点以后根轨迹偏离了虚轴。而且，未加考虑与建模的寄生元件也使得结果变得不可靠。而允许有为这些因素影响的振荡频率不是我们期望的。然而，尽管有着这些障碍，有报道实现了性能良好的正交相位(误差小于 0.5°)输出的 1GHz 振荡器[24]。

15.6　负阻振荡器

一个完全无损的谐振电路非常接近于一个振荡器，但是无损耗的元件是很难实现的。要实现实际的振荡器，一个可能的方法是用有源器件的能量供给能力来补充实际谐振回路中由于有限的 Q 值而损耗的能量，就像 TITO 这个例子。

前面的描述是相当一般的，包括了反馈和开环的拓扑结构。属于前者的是一个经典教科书上的电路，即负阻抗变换器(NIC)。NIC 可以通过在一个集成运放电路上同时加正反

---

[24] R. Duncan *et al*.. A 1-GHz Quadrature Sinusoidal Oscillator(一个 1 GHz 正交正弦振荡器). *IEEE CICC Digest*, 1995, pp. 91-94.

馈和负反馈来实现。特别是考虑图 15.29 中给出的电路。

假设集成运放的特性是理想的，很容易就可以得到输入阻抗与反馈阻抗的关系为

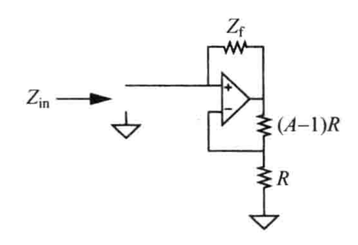
$$Z_{in} = \frac{Z_f}{1-A} \tag{40}$$

如果闭环增益 A 的值被恰好设置为 2，那么输入阻抗的值将是反馈阻抗的负值。如果环路中的反馈电阻选择为有正值的纯电阻，那么输入电阻就将是纯负的电阻。这个负电阻可以用来补偿所有实际的谐振器中的正电阻来得到振荡器。

通常来说，所有真实的有源器件的本征的非线性性会对幅度有一个限制，而且如果希望的话，可以用描述函数来估计振荡的幅度。描述函数也可以被用来确定振荡器是否事实上会振荡。

作为一个专门的例子，考虑图 15.30 所示的振荡器。为了保证振荡，需要跨越谐振回路两端的纯电阻为负值。所以必须满足下边的不等式：

$$R_t > R_f \tag{41}$$



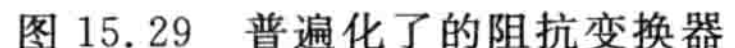


图 15.29　普遍化了的阻抗变换器

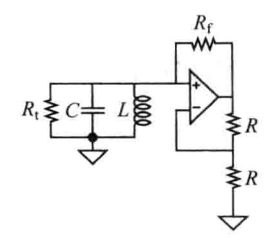


图 15.30　负电阻振荡器

在低频情况下，限制幅度的最为典型的非线性是实际放大器有限输出摆幅。因为从谐振器到集成运放输出的增益为 2，所以谐振回路两端上的信号范围被限制到了略高于电源的一半，这对应于放大器输出的周期性饱和。

在更高的频率上，可能会发生有限的放大器摆率起到幅度的控制作用（这是部分的，非全部）。通常并不希望这种情况出现，因为与有限的摆率有关的相位延迟会造成振荡频率的漂移。在一些极端的情况下，摆率对幅度的限制（或者是放大器的其他形式的幅度限制）是不稳定的，因而有可能引起间歇振荡。

最后提一下，先前所提到的各种振荡器（例如，Colpitts，Pierce 等）都可以看成是负阻振荡器。

一个更加实用的负电阻可以通过利用寄生效应容易得到：在一个共基放大器的基极电路中的电感能够在发射极端出现一个负电阻，参见图 15.31。一个直接的分析揭示了对频率高于由这个电感与 C_{be}（假若 C_{cb} 可以忽略）组成的回路的谐振频率，Z_{in} 的实部是负的。对于高于谐振回路的谐振频率而远小于 ω_T 的频率，Z_{in} 的实部近似为

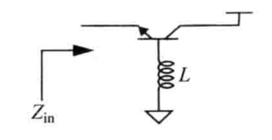


图 15.31　规范化的射频负阻（未标出偏置电路）

$$R_{in} \approx -\frac{\omega^2 L}{\omega_T} = -\frac{\omega}{\omega_T}|Z_L| \tag{42}$$

因为这个电路可以很容易地实现一个负电阻，所以它的应用是很广泛的。但是也要注

意，当我们不需要一个负阻的时候，保持尽可能小的基极寄生电感是很重要的。

在最近这几年，一个用到了交叉耦合的差分对来综合产生负阻的电路经常被用到。对这个电路的分析作为练习留给读者，如图 15.32 所示。

在第 17 章中将会讲到，使信号幅度最大化可以提高信号的频谱纯度(因为信噪比的提高)。在类似图 15.32 所示的许多振荡器结构中，所允许的最大信号幅度是由电源电压或器件的击穿电压所决定的。因为信号是由在谐振回路中所存储的能量决定的，所以也就可以从 Clapp 振荡器中得到启示，采用中间抽头的谐振回路来使谐振回路的峰值幅度超过器件的击穿电压或者电源电压，就像是在负阻振荡器[25]一样，见图 15.33[26]。

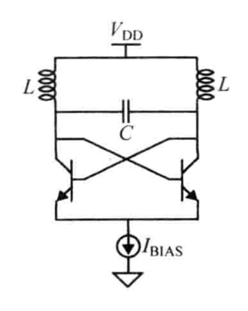


图 15.32　简单差分负阻振荡器

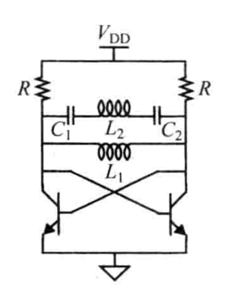


图 15.33　带有修改过的谐振槽路的负阻振荡器(简化版本)

差分的连接方式使得我们不太容易看清用到了中间抽头的共振器，所以为了简单起见，仅仅考虑一半的电路(见图 15.34)。在这个简化的半电路中，晶体管被一个负阻代替了，正的电阻则根本没有表示出来。而且，那两个电容被它们的串连等效电容代替了，而电感代表的是原始电路中漏极连接的作用。

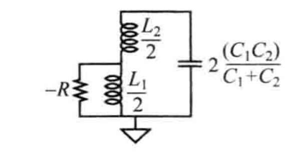


图 15.34　负阻振荡器的简化了的半电路

很明显的，由于用到了抽头结构，等效电容(或者等效电感 L_2)上的电压幅度可以超过电源电压(甚至超过器件的击穿电压)。因此，这种振荡器在思想上与 Clapp 结构是相近的。实际用到的输出信号可以从插入在两个振荡器核心中间的任何一个缓冲器上得到，否则一个容性的分压会毁坏共振器的 Q 值。作为一个或者两个抽头的共振器中所存储的高能量的结果，这个电路具有非常优秀的相位噪声性能，这些在下一章中加以讨论。

要调谐这个(或者其他 LC 的)振荡器，可以将全部的或者部分的电容选作可变电容(例如，n 阱中的 p+掺杂形成的节电容)，并精确地调节偏置控制电压以使得它的等效容达到谐振状态。因为 CMOS 的节电容的 Q 值是十分糟糕的，所以强烈建议只使用必要的节电容来达到必须的谐振范围就可以了。在实际应用中，如果忽略过多的相位噪声的蜕化作用，调谐的频率范围被限制在低于 5%～10%。图 17.39 中显示了一个应用这种结构的电压控制振

---

[25] J. Craninckx and M. Steyaert. A CMOS 1.8GHz Low-Phase-Noise Voltage-Controlled Oscillator with Prescaler (一个带有预分频器的 CMOS 1.8GHz 低噪声压控振荡器). *ISSCC Digest of Technical Papers*, February 1995, pp. 266-267. 电感是跨接在管芯上的引线电感。

[26] 最好是有一个尾电流源来抑制摆幅，但是为了简化起见，忽略了这一细节。

荡器，很简单但是有示范意义。

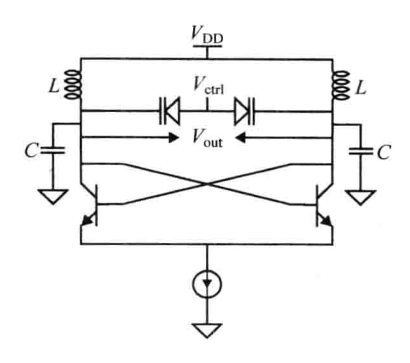


图 15.35　压控负阻振荡器(简化版本)

作为负阻振荡器讨论的最后一个注解，必须清楚许多(即使不是全部)振荡器都可以被看作负阻振荡器。因为从谐振器的角度来看，有源器件抵消了非无穷大的谐振槽路 Q 值所造成的能量损失。因此，是否称一个振荡器为"负阻"振荡器更多的是哲学上的决定，而不是本质上的不同。

15.7　总结

本章中考察了振荡器的振幅是怎样通过非线性的手段建立起来的以及将额反馈的概念扩展以包括一个特别类型的线性化非线性：描述函数。有了描述函数及环路传输链是其他元件的知识就可以确定振荡器的频率和振幅。

我们讨论各种振荡器，包括开环和反馈结构的振荡器。Colpitts 和 Hartley 振荡器采用抽头谐振腔来提供正反馈，TITO 振荡器采用调谐放大器加上米勒反馈提供的负阻实现振荡。Clapp 振荡器采用了一个额外的抽头来提供超出电源电压的摆幅，这使得信号能量能大大高过噪声能量。

我们还给出了基于晶体振荡器的 LC 振荡器。由于石英晶体表现十分像一个具有非常高 Q 值的 LC 谐振器，人们可以用它来实现非常高的频率纯度以及非常低的功耗。Colpitts 结构的振荡器工作在石英的串联谐振频率，因此需要一个 LC 谐振器。Pierce 振荡器工作在石英呈现为电感时的频率，从而不需要额外的电感。然而，由于工作在偏离谐振状态，因此迫使对于特定的负载电容需要特制一个石英晶体。

我们也随机考察了一些其他的振荡器，包括在环路中采用两个积分器的正交振荡器以及几种不同的负阻振荡器。我们又一次可以看出，抽头谐振器对于改善相位噪声很有益处。

第16章 频率综合器

16.1 引言

在现代通信系统中，锁相环(PLL)已经变成无处不在，因为它具有突出的通用性。作为一个重要的例子，锁相环可以用来产生一个输出信号，其频率是可编程的，等于一个固定输入频率的有理分数倍。这种频率综合器可用作超外差收发机的本地振荡信号。锁相环也可被用来进行频率调制和解调，以及从一个载频已经被压制的输入信号中重新产生载频。锁相环的通用性也涵盖了纯数字电路系统，在那里它在信号扭斜(skew)补偿、时钟信号恢复和时钟信号产生都是不可或缺的。

为了详细地理解 PLL 是怎样能实现如此多样的功能，需要建立这些反馈系统的线性模型。当然，首先简单地回顾一下历史以便把本章的议题置于恰当的上下文中。

16.2 锁相环(PLL)简史

现在被称为 PLL 的最早描述是 H. de Bellescize 于 1932 给出的①。这个早期的工作采用了中频为零的超外差接收机的再生方式提供了一个接收和解调调幅(AM)信号的另一个结构。用这种零中频选择，就不需要排除镜像频率。而且所有的向下的频率转换(下变频)处理都在音频范围内进行。

然而，要使零差或直接变频接收器正确地工作要求本地振荡器(LO) 的频率和输入的载频完全相同。更进一步说，为了达到最大的输出，LO 必须和输入的载频同相位。如果相位的关系不能被控制，增益会很小，甚至为零(比如 LO 正好和载波正交)，或者增益的大小变化剧烈。De Bellescize 提出了一种解决该问题的方法，就是将 LO 的相位锁定到载波的相位。

因为各种各样的原因，零差接收器没有取代通常的超外差接收器。后者在大约 1930 年之后就在无线电市场上占主要地位。然而，由于零差结构对(镜像频率)滤波的要求较低，因此有利于单片集成，近年来，人们对这种零差接收器又重新感兴趣起来②。

另一个类似于 PLL 的电路用在电视上已经 30 多年了。在标准的电视节目传送中，两个锯齿波发生器分别产生一个垂直的和一个水平的偏转(扫描)信号。为了使接收器能够和电视台发出的扫描信号同步，在发送音频和视频信号的同时要发送一个同步脉冲信号。

---

① La Reception Synchrone. L' Onde Électrique. v. 11, June 1932, pp. 230-240.

② 然而零差对接收前端线性度和直流偏差的容忍性要求很高。同时因为 RF 和 LO 的频率是一样的，LO 信号从天线反向漏泄出去是一个问题。再有一个问题是，这个 LO 泄漏还会返回到前端来，和 LO 信号以随机的相位混频。其结果是产生一个变值的 DC 偏差，它的值可以比 RF 信号大几个数量级。这些问题或许象滤波问题那样难以解决，我们将在第 18 章中更详细地讨论。

早期的电视机里同步是这样进行的：电视机的扫描振荡器被调节到一个比实际的传输频率低的频率自由震荡着，用一种叫做“切入锁定”的技术[③]，让同步脉冲使锯齿波振荡器提前结束每一个周期，从而实现同步操作。只要收到的信号拥有相对比较小的噪声，这个操作工作得相当好。然而，随着信噪比变小，同步操作越来越不能正常工作，要么是同步信号丢失，要么是噪声被误当作同步脉冲。这种电路广泛应用的时候，每一个电视机都有一个垂直同步旋钮和一个水平同步旋钮，这样观众可以自己摆弄着两个旋钮来调节接收器自由震荡频率，保证仍能达到同步。如果那两个旋钮调节不当，会使电视画面上下或者左右翻滚。现在，即使信号的信噪比很差，电视机里面的 PLL 电路仍然能够很好地提取出同步信息，因此人们再也用不着这两个垂直同步和水平同步的旋钮了。

还有一个类似 PLL 的电路还是用在电视上。早在 19 世纪 40 年代末期和 50 年代初期。当彩色电视的概念开始被引入的时候，联邦无线委员会(FCC)提出了一个和当时黑白电视标准相兼容的要求，并声称彩色电视不应该占据额外的带宽。因为黑白电视发明的当时并没有考虑后来的彩色电视的要求，因此要满足上述限制显然并不是一件小事情。讲得细致一点，就是怎样才能把彩色电视信号压入到和黑白电视同样的带宽中去，而不丢失其他任何信息？问题的突破点是人们注意到电视的 30 Hz 的帧信号频谱不是连续的，而是梳状的，每隔 30 Hz 有一个尖峰。色彩信息就可以嵌入到这些尖峰间的空隙中，而不占据额外的带宽。为了利用这一点，附加的色彩信息被调制在一个大约频率是 3.58 MHz 的子载波上[④]。仔细挑选子载波的频率使得色度信号的频带边沿正好在黑白信号频谱尖峰间的中点位置下降为零。复合的黑白信号(又叫“亮度”信号)和色彩信号然后再去调制最后的载波进行发送。这是美国的制式，叫做 NTSC(National Television Standards Committee)制式。

色彩信息是矢量编码的，其相对于子载波的相位决定了色调，而它的幅度决定了颜色的强度(色饱和度)。接收器必须能够提取或者产生一个子载波来产生一个 0°的相位参考，否则，还原出来的颜色就与原来的颜色不一样了。

为了实现这种锁相，电子束从屏幕右边回扫到左边的这段时间里，视频信号里要出现一串 3.58 MHz 的参考脉冲信号(NTSC 标准规定至少 8 个脉冲)。这一串脉冲信号馈入接收器中专门再产生 3.58 MHz 连续子载波信号的振荡器里，从而使振荡器和这一串信号的相位锁住。因为这个脉冲流不是连续的，当电子束在屏幕上正程扫描图像时，没有这种参考脉冲，这时振荡器必须能够自己维持连续的振荡，并且这段时间内，再生子载波的相位不可以漂移以避免色彩的漂移。早期的技术并不能完全实现这个目标，以至于 NTSC 一度被戏称为“永远不可能重现色彩”(Never twice the same color)[⑤]。

---

③ 参考 Balth. van der Pol. Forced Oscillations in a Circuit with Nonlinear Resistance (Reception with Reactive Triode). *Philosophical Magazine*, v. 3, January 1927, pp. 65-80，以及 R. B. Adler, “A Study of Locking Phenomena in Oscillators”, *Proc. IRE*, v. 34, June 1946, pp. 351-357. 人体的 24 小时生物钟是切入锁定的另一个例子。如果没有太阳光(或者其他周期光源)的同步，对于大多数人来说一“天”实际上是 25～27h。可见自由震荡频率又比同步的频率低一些。

④ 更精确的说，是 3.579 545MHz，用视频和音频载波的差 4.5MHz 再乘以因子 455/572 得到的。

⑤ 实际上，非常非常早的时候，这种电路甚至不用振荡器，而是拿那串参考脉冲去激励一个很高 Q 值的谐振回路(一般用石英晶体)，参考脉冲过后，谐振回路持续的振荡波形被用作子载波，并且要持续大约 200 个周期而没有过分的衰减。

欧洲(除了法国[⑥])采用了一个类似的色度编码机制,只不过在电子束每扫过一条线时变换一次电压的参考极性,这样在电子束相继两次扫描后,相位漂移平均起来就为零,这种方法大大减少甚至消除了颜色的漂移,因此叫做 PAL 制式(Phase Alternating Line)。

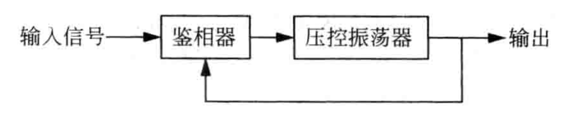


图 16.1 锁相环结构

另一个类似 PLL 电路的早期应用是在调频(FM)立体声广播上。同样是为了向后兼容性,立体声的信息是调制在 38kHz 的子载波上的。把左右两个声道求和(L+R)作为单声道信号(限制带宽为 15kHz),把两声道的差分信号(L-R)调制在 38kHz 的子载波上然后进行广播。这个差分信号经调制后变成双边带抑制载波信号(DSB-SC)。接收器会重现这个 38kHz 的子载波,并且把单声道信号(L+R)和差分信号(L-R)经过简单加法和减法复原出原来的左声道信号和右声道信号。为了简化接收器的设计,发送的信号中包含一个小幅度的"引导"信号,其频率恰好是子载波频率的一半。接收器把这个引导信号倍频以后去解调出 L-R 信号来。很快我们可以看到,即使没有这个引导信号,一个 PLL 也能够轻而易举地实现这个倍频操作。但在 1960 年,这个引导信号是对电路非常大的帮助了。

早期的 PLL 都是"注入锁定"类型的电路或者它的变种。由于电路太复杂,教科书上的 PLL 电路对于大多数消费类应用来说都太庞大而不适用,除非用在特殊的场合,例如卫星通信或者科学仪器上,而且数量也不多。直到 20 世纪 70 年代,IC 技术的迅速发展使得 PLL 电路可以真正用在 FM 解调中,从此,PLL 变得越来越广泛,从普通的系统到高度专业化的系统都有应用。

由此,我们知道了锁相技术具有广泛的应用。以此背景知识为动机,接下来研究"教科书"中的 PLL 的建模问题。

16.3 线性化的锁相环模型

基本的 PLL 结构如图 16.1 所示,由一个鉴相器和一个压控振荡器(VCO)构成[⑦]。鉴相器把输入信号的相位和 VCO 输出信号的相位作比较,并生成一个与相位差具有某种函数关系的输出电压。而 VCO 只是简单地产生一个振荡频率和控制电压有某种函数关系的周期信号。

PLL 的基本思想是用鉴相器的输出去驱动 VCO 频率,使得相位差减小。换句话说,这是一个负反馈系统。理论上,一旦这个反馈环锁住以后,输入参考信号和 VCO 的输出信号之间就有一个固定的相位关系(通常是 0°或者 90°,与鉴相器的性质有关)。

尽管实际的鉴相器和 VCO 有高度的非线性,通常的做法是在相位已经锁住的时候假

⑥ 法国的彩色电视系统是 SECAM 制式,Sequentiel Couleur avec Memoire。这个系统的亮度信号和色彩信号是分时传送的,在接收器端再把它们分别重组起来。

⑦ 一些振荡器是通过调节电流来控制频率的。但习惯做法是将电流和电压控制振荡器都称作 VCO,除非有必要进一步区分二者。

设为线性的。以后会考虑一个更加一般的情形(包括捕获过程)，但是现在先从最简单的开始，然后逐步增加复杂度。

那么让我们来看图 16.2 所示的线性化的 PLL 模型，因为我们更加关心输入和输出之间的相位关系，因此在这个模型里面输入变量和输出变量都是相位，而不是它们实际上的时域波形。如果你习惯把框图里的“信号”理解为电压，那么现在请把信号理解为相位。

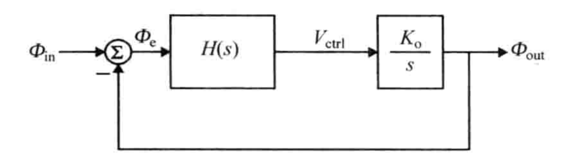


图 16.2　线性化 PLL 模型

把相位作为输入-输出变量的另一个影响是，输出频率取决于其控制电压的 VCO，被当作一个对频率的积分器来处理，因为相位是频率的积分。VCO 的增益常数 K_O 的单位是：rad/(s·V)，仅仅用来描述当控制电压发生一个已知的改变时，输出频率会发生什么样的变化。请注意，与通常输出有界的放大器不同，这里的 VCO 真的是一个积分器，等待时间越长，积累的相位越大(除非有人将 VCO 关掉)。

鉴相器被当作一个简单的减法器来处理，产生一个输入信号和输出信号相位误差 Φ_e。为了表示增益比例因子和环路中可能有的滤波效应，在模型中采用了一个具有传输函数 H(s)的模块。

16.3.1　一阶 PLL

最简单的 PLL 的传输函数 $H(s)$ 是一个简单的常量增益(记做 K_D，单位：V/rad)。因为这个环路只有一个极点，因此叫做一阶 PLL。除了简单，一阶 PLL 还能够轻而易举地获得很大的相位裕度。

但是一阶 PLL 有个大的缺点就是带宽和稳态的相位误差有强烈的耦合。因为通常人们希望稳态的相位误差和带宽没有关系，因此一阶 PLL 很少有人用。

现在用一阶线性 PLL 模型来定量地分析一下一阶 PLL 的限制。其输入和输出的相位传输函数由以下式子给出：

$$\frac{\Phi_{out}(s)}{\Phi_{in}(s)} = \frac{K_0 K_D}{s + K_0 K_D} \tag{1}$$

于是闭环带宽为

$$\omega_h = K_0 K_D \tag{2}$$

为了验证带宽和相位误差是相互关联的，我们来看输入-误差传输函数：

$$\frac{\Phi_e(s)}{\Phi_{in}(s)} = \frac{s}{s + K_0 K_D} \tag{3}$$

如果假设输入信号是一个固定频率 ω_i 的正弦信号，那么其相位是以一个 ω_i(rad/s)的速度线性增长的。因此输入信号在拉普拉斯(Laplace)域的表达式是

$$\Phi_{in}(s) = \frac{\omega_i}{s^2} \tag{4}$$

这样

$$\Phi_{e}(s) = \frac{\omega_{i}}{s(s + K_{0}K_{D})} \tag{5}$$

于是对于固定输入频率的稳态误差为

$$\lim_{s\to 0} s\Phi_{e}(s) = \frac{\omega_{i}}{K_{0}K_{D}} = \frac{\omega_{i}}{\omega_{h}} \tag{6}$$

由此可见，稳态相位误差仅是输入频率与环路带宽之比，当环路带宽等于输入频率时，得到一个单位弧度的相位误差。因此，当需要小的稳态相位误差时，环路带宽就要比较大，正如前面所指出的那样，这两个参数是紧紧联系在一起的。

有一种直观的方法可以对以上结果进行解释就是：一般来说，VCO 都需要一个非零的控制电压来校正它的输出频率，既然这个控制电压是由鉴相器输出的，那么就必然会有一个非零的相位误差。对于一个固定的输出如果要减小相位误差，必然要求鉴相器有一个很大的输出增益。因为增益的增加对环路传输函数的影响在各个频率点上是一样的，因此带宽的增长必然包含着相位误差的减小。

为了实现零相位误差，需要一个能够从鉴相器的零输出产生出一个任意控制电压的元件，这意味着需要一个具有无限增益的元件。为了使稳态相位误差和带宽没有关系，这个元件应该在 DC 的时候具有无限增益，而不是在所有频点。积分器就有这个功能，使用了积分器的 PLL 就是二阶 PLL。

16.3.2　二阶 PLL

二阶 PLL 的模型如图 16.3 所示。积分器引入的 $-90°$ 相移必须被一个环路稳定的零点的正相移补偿回来。对于这种方式的反馈系统的补偿，上述零点的位置必须比交叉频率的位置低很多从而获得期望的相位裕度。

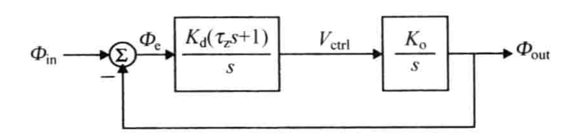


图 16.3　二阶 PLL 模型

在这个模型里面，因为额外的积分环节，常数 K_D 的单位变成 rad/s。同样归功于这个积分环节，PLL 环路带宽就和稳态相位误差(在这里是零)再也没有关系了，这一点如果我们分析图 16.4 的环路传输函数幅度就可以看得很清楚。用二阶 PLL 的根轨迹(图 16.5)可以分析它的稳定性。

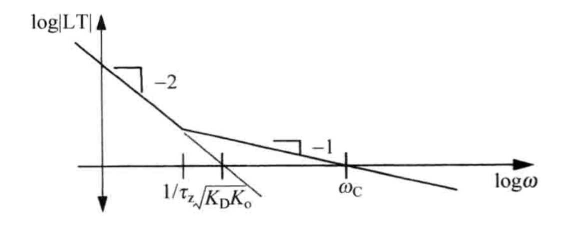


图 16.4　二阶 PLL 的环路传输系数

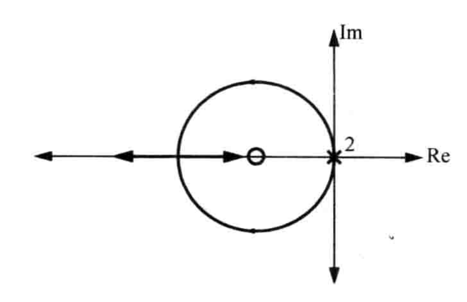


图 16.5　二阶 PLL 的根轨迹图

增加环路传输函数的幅度(通过增加 $K_D K_0$)，由于交叉频率的增加会使得零点的正的相移能够更多地补偿到极点的负的相移中去，环路衰减会变得更好。当环路传输系数幅度非常大时，其中一个闭环极点位于频率为零处附近，另一个极点则趋向无穷。

在这种 PLL 里，环路稳定零点来自前馈的那条支路，因此它也将出现在闭环的传输函数中。

容易得到相位传输函数：

$$\frac{\Phi_{out}}{\Phi_{in}} = \frac{\tau_z s + 1}{(s^2/K_D K_0) + \tau_z s + 1} \tag{7}$$

则

$$\omega_n = \sqrt{K_D K_0} \tag{8}$$

和

$$\zeta = \frac{\omega_n \tau_z}{2} = \frac{\tau_z \sqrt{K_D K_0}}{2} \tag{9}$$

并且环路的交叉频率为

$$\omega_c = \left[\frac{\omega_n^4}{2\omega_z^2} + \omega_n^2 \sqrt{\frac{1}{4}\left(\frac{\omega_n}{\omega_z}\right)^4 + 1}\right]^{1/2} \tag{10}$$

通常，交叉频率远大于零点的频率，于是式(10)可以简化为

$$\omega_c \approx \frac{\omega_n^2}{\omega_z} \tag{11}$$

图 16.4 和式(10)都说明交叉频率总是比 ω_n 大，而由图 16.4 及式(11)可知，ω_n 是环路没有零点时的外推交叉频率。最后要指出的是，给定 ω_n 以后，增加零点的时间常数能够增加环路衰减。这样，在保留稳态零相位误差的条件下，二阶 PLL 电路的带宽和稳定性都可以调节到事先的要求上。

16.4　输入端噪声的 PLL 抑制

我们看到，使 PLL 的带宽最大化能够减少 VCO 频率扰动所带来的影响。这个观点并不深奥：不管什么对于信号源，一个快速系统意味着它能够更快地从误差中恢复过来。

然而，最大化带宽有一个比稳定性更加严重的潜在问题。随着环路带宽增加，它能够更好地追踪输入信号。如果输入没有噪声(或者至少比 PLL 自己的噪声少)，那么总的效果是

好的。但是，如果输入信号比PLL的VCO的噪声更多，那么这个高带宽的环路会忠实地把输入噪声信号在输出端重现出来。因此，在对于环路输入端噪声的敏感性和对于干扰VCO频率的噪声的敏感性这两者之间就有一个折中，前者希望环路带宽要小，后者则希望环路带宽越大越好。

一般来说，对于一定的功耗，调谐振荡器(例如LC振荡器或者基于晶体的)比延时振荡器(例如环形振荡器或者RC移相振荡器)具有更少的噪声。因此，如果进入PLL的参考输入信号是由一个调谐振荡器产生的，而VCO却是基于延时振荡器的拓扑，那么我们更希望有大的带宽。如果情况刚好相反(一般少见)，延时振荡器提供一个输入参考信号给一个基于晶体振荡器的PLL，那么一般希望环路带宽要小。

16.5 鉴相器

我们已经在框图层次上了解了传统的PLL，并且特别注意了二阶PLL锁住时的线性化行为。现在要考虑一些实际的细节来看看真正的PLL是怎样构建以及工作的。在这一节里，我们会看到几个有代表性的鉴相器。

16.5.1 模拟信号乘法器作为鉴相器

在输入是正弦波并且VCO也是正弦波的PLL里面，迄今最常用的鉴相器就是乘法器，一般用吉尔伯特(Gilbert)类型的拓扑来实现。对于一个理想的乘法器来说，要得到输入-输出关系并不困难(如图16.6所示)。

用一些三角恒等式，我们发现乘法器的输出可以表示为

$$AB\cos\omega t\cos(\omega t+\Phi)=\frac{AB}{2}[\cos\Phi-\cos(2\omega t+\Phi)] \tag{12}$$

注意，乘法器的输出包含一个直流项和一个倍频项。对于鉴相器来说，我们只关心它的直流项，这样，鉴相器的平均输出为

$$\langle AB\cos\omega t\cos(\omega t+\Phi)\rangle=\frac{AB}{2}\cos\Phi \tag{13}$$

可见鉴相器的增益"常数"是相角的函数，由下式给出：

$$K_{\mathrm{D}}=\frac{\mathrm{d}}{\mathrm{d}\Phi}\langle V_{\mathrm{out}}\rangle=-\frac{AB}{2}\sin\Phi \tag{14}$$

如果把鉴相器的平均输出作为相角的函数画在图上，那么可以大概得到图16.7的那样一条曲线。注意，鉴相器输出具有周期性。进一步观察又可以发现，当相差是零时鉴相器的增益也是零，而当输入相差为90°时增益最大。这样为了最大程度利用鉴相器的输出，环路应该被安排在相差为90°时锁住。因此，乘法器又常常被称作"正交"鉴相器。

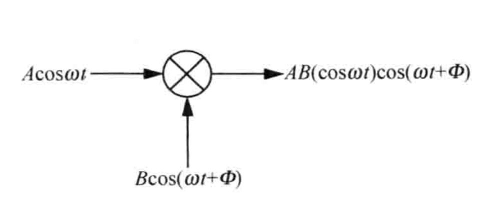


图16.6 乘法器作为鉴相器

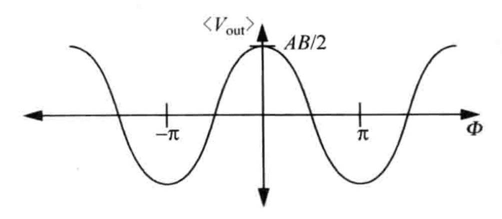


图16.7 乘法器鉴相器的输出与相位差的关系

当环路被锁在正交状态时，鉴相器就有一个递增的增益常数：

$$K_D\big|_{\phi=\pi/2}=\frac{\mathrm{d}}{\mathrm{d}\phi}\langle V_{out}\rangle\bigg|_{\phi=\pi/2}=-\frac{AB}{2} \tag{15}$$

下面我们会自然而然地忽略那个负号，这是因为环路应该伺服在相差等于 90°或者 −90°的状态（当然不是同时达到这两个角度），负号取决于环路中其他部分的反相次数。

在一个 2π 的周期内有两个角度会使得鉴相器的输出为零。这样就有两个平衡点鉴相器能够锁住。但是，其中一个点是稳态平衡，而另外一个点是亚稳态平衡，从这个点环路最终是要发散的。换句话说，只有其中一个点对应着负反馈。

当提到正交 PLL 环路的相位误差时，我们计算的是与 90°相位差的平衡条件的偏离程度，因此，尽管在一个理想的正交 PLL 环路中相差为 90°，但是我们说相位误差为零。

16.5.2 可互换的乘法器作为鉴相器

在前一节里，我们假定环路的两个输入都是正弦波。然而，在大多数实际应用中，其中一个信号很可能是近似的方波。那么我们现在把输入改成一个方波，这样得到如图 16.8 所示的框图，其中“sgn”是符号函数，定义为：

$$\mathrm{sgn}(x)=1,\quad x>0 \tag{16}$$

$$\mathrm{sgn}(x)=-1,\quad x<0 \tag{17}$$

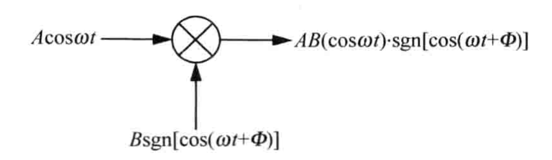


图 16.8 带有一个方波输入的乘法器

现在，回忆一下一个幅度为 B 的方波，它的基波分量的幅度就是 $4B/\pi$。如果只关心基波分量，那么乘法器的平均输出为

$$\langle V_{out}\rangle=\frac{4}{\pi}\frac{AB}{2}\cos\Phi=\frac{2}{\pi}AB\cos\Phi \tag{18}$$

于是相应的鉴相器的增益就是上一节纯粹的正弦波的 $4/\pi$ 倍：

$$K_D\big|_{\Phi=\pi/2}=\frac{\mathrm{d}}{\mathrm{d}\Phi}\langle V_{out}\rangle\bigg|_{\Phi=\pi/2}=-\frac{2AB}{\pi} \tag{19}$$

尽管鉴相器的输出和增益对于方波和纯粹的正弦波来说非常类似，但是这两类的鉴相器还是有性质上的重要差别。因为方波除了基波分量还拥有其他分量，环路可能会锁在输入信号频率的谐波或者分频波上。例如，考虑频率刚好是正弦波 1/3 的一个方波，这样方波⑧只包含有奇次谐波，而第三次谐波的频率就正好和正弦波相同了，这两个信号都会让乘法器产生一个直流输出。

⑧ 我们隐含假设方波的占空比是 50%，非对称方波除了含有奇次谐波之外也含有偶次谐波，有可能使得环路除了锁在输入参考信号的奇次谐波以外，还可能锁在偶次谐波上。

因为方波的频谱以 $1/f$ 的速度衰减[9]，越来越高阶的谐波的平均输出自然是越来越小。同时，鉴相器增益常数的减小又使得环路锁住并维持在高次谐波上更加困难，但是只要是使用此类鉴相器的实际环路都还是要考虑这个因素。有的时候锁频在谐波上正是我们所需要的，有的时候不是。如果不是，为了防止锁住在谐波上，则通常 VCO 的频率范围要受到限制（或者要仔细协调好）。

另一个值得注意的地方就是，一个信号乘以一个周期符号函数等效于把这个信号的相位周期性地颠倒一下。这样，这样被使用的乘法器就可以被一些开关来代替（所以也叫做“换向器”，类似于转动机械里面的一个装置）。无源二极管环形混频器可用作换向混频器。在有些 IC 工艺里（例如 CMOS），换向混频器是对 Gilbert 乘法器的补充。

16.5.3　异或门作为鉴相器

如果驱动模拟乘法器输入的两个信号都是方波，那么可以把每一个输入展开为傅里叶级数，把它们相乘这样来处理。然而，看起来在时域分析这样一种特殊的情况更为容易一些，因此将按照这个思路来进行分析。读者可以把在频域的分析作为另外的练习做一下，并比较一下这两种方法（鼓励这么做）。

现在，两个输入都是方波产生的输出如图 16.9 所示。当我们改变相差时，输出的形式仍然是一个方波，不过占空比会改变，如果输入信号正交的话，那么输出刚好是 50%的占空比。因为输出占空比跟输入信号的相差成正比，就可以容易地把输出的平均值和输入相差之间的函数关系画一个图（如图 16.10 所示）。

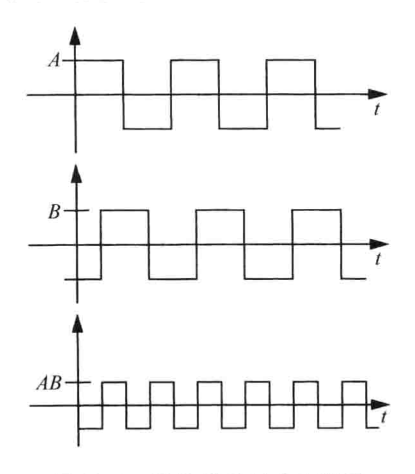


图 16.9　乘法器的输入与输出

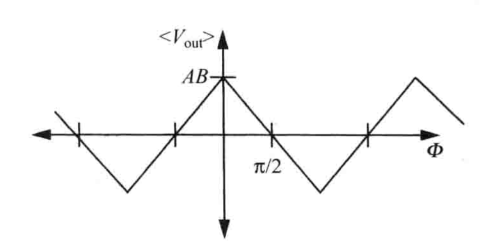


图 16.10　带有两个方波输入的乘法器特性

现在这种情况下，鉴相器的增益是一个常数，等于

$$K_D = \frac{2}{\pi}AB \tag{20}$$

我们看到，这个值与输入信号为纯粹的正弦波并且正交的时候完全一样，只是多了一个

⑨　一个让人有点迷惑的有趣的事情是，一般来说，信号的频谱是以 $1/f^n$ 的速度衰减，其中 n 是让这个信号变成一个冲击函数需要求导的次数。于是，理想正弦波的频谱的降落是很快的（因为无论求导多少次，正弦波都不会变成冲击函数），而冲击函数频谱永远不会衰减（因为 $n=0$），方波的频谱以 $1/f$ 的速度衰减，三角波频谱衰减速度是 $1/f^2$，等等。

比例因子。

在其中一个输入是方波的情况下，这个鉴相器允许环路锁定在某个谐波上。当然，要根据实际情况应用，这种特性可能是或者不是我们所希望的。

如果更加仔细地检查一下鉴相器的输出波形，会发现正好和数字电路异或门的情况完全一样，唯一的不同是直流的偏置不一样，还有相位反相的位置。因此，一个异或门可以被当作一个过驱动的模拟乘法器来考虑。更特殊的情况，当输入和输出的信号是逻辑电平，在地和某个电源电压 V_{DD} 之间摆动时（比如在 CMOS 中），鉴相器的输出平均值就如图 16.11 所示。相应的鉴相器增益是

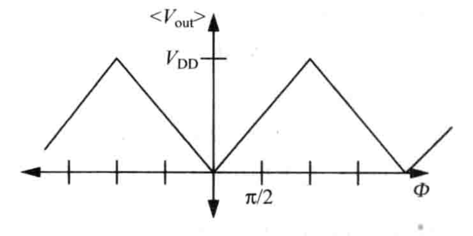


图 16.11　作为正交鉴相器的异或门的特性

$$K_D = \frac{V_{DD}}{\pi} \tag{21}$$

异或门鉴相器很容易实现，并且和其他数字电路兼容，因而在简单的集成 PLL 电路里经常被采用。

16.6　序列鉴相器

采用乘法器作为鉴相器的锁相环是在输入信号之间呈正交相位关系的时候锁住的。但是，有一些实际的情况是需要在零相差的时候锁住（de Bellescize 的零差 AM 检波器就是一个例子）。并且，鉴相器在所期望的平衡点和亚稳态点上的增益大小相同，这意味着它会在亚稳态停留很长的时间，可能导致达到锁住的时间变长。

序列鉴相器能够供在零相差（或者是 180°）锁相，并且它的平衡点和亚稳态点的增益差很多。而且，一些序列鉴相器的输出和相位误差成正比的范围甚至能够超过 2π。

当然，序列鉴相器确实有一些缺点，因为它们工作在跳变状态，因此对边沿的丢失很敏感（尽管可以做一些修正来减小这种敏感性），这是和乘法器截然不同的，后者看到的是波形的全部。而且，这个边沿触发特性带来的另一个后果就是它们对环路引入了一个采样操作。以后我们会看到，采样操作会对环路传输函数增加一个类似于延时的效应。随着频率而增加的负的相移给所能允许的交叉频率又增加了一个上限，这使得它比起其他鉴相器来，又增加了一个很大的限制。

16.6.1　具有增宽输入范围的序列鉴相器

具有上述特性的一个用得较广泛的电路是用两个 D 触发器和一个复位门构成的（如图 16.12），符号“R”和“V”分别表示“参考信号”和“VCO 信号”，而“U”和“D”则分别表示“上”和“下”，是指短暂发生的事件。

这个电路的上、下两个输出端的差分电压的平均值和输入相差的关系如图 16.13 所示。可见，现在的输入范围扩展到了 4π，且常数鉴相增益为

$$K_D = \frac{V_{DD}}{2\pi} \tag{22}$$

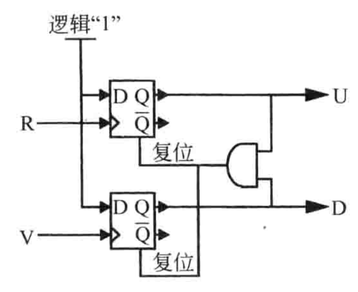
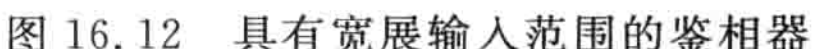


图 16.12 具有宽展输入范围的鉴相器



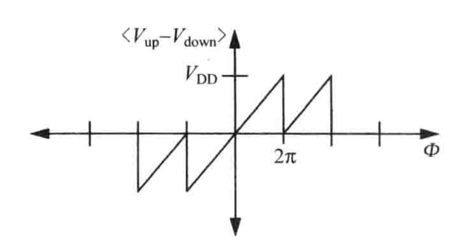


图 16.13 具有宽展输入范围的鉴相器特性

偶尔引起问题的一个特性是可能产生的矮脉冲。如果图 16.12 中的 reset 路径动作太快,U 和 D 输出的最小脉冲宽度对于下一级的可靠工作可能过窄。当 R 和 V 输入非常近时,该问题就会出现,这会使锁相点处的特性恶化。这种恶化一般会以无法可靠地解决靠近锁相点处的相位误差的形式出现。通过使 reset 路径变慢可以很容易地解决这一"盲区"问题。在与门之后插入恰当数量的反相器就可保证 U 和 D 的输出波形宽度一致,能够使后续各级正常工作。锁相后,对于相同的时间间隔,U 和 D 的输出都是有效的。

16.6.2 鉴相器与鉴频器的比较

在很多的应用中,知道两个输入信号之间的频率差的大小往往很重要(或者至少很有用)。例如,该信息有助于锁相捕获。

基于乘法器的鉴相器不能给出这样的信息,而序列鉴相器却可以。考虑上面提到的扩展输入范围的鉴相器,如果 VCO 的频率比参考信号的高,那么 U 的输出将会呈现高占空比,因为它是由频率高的 VCO 的上升沿来置位的,但是直到频率低的参考信号的上升沿到来才复位。因此,这种类型的鉴相器不但可以提供一个宽的和线性的鉴相范围,而且能够提供一个关于频率误差的符号和大小的信息。这些性质使得这种鉴相器大受欢迎。能够区分频率差别的鉴相器通常统称为"相频"检测器。

当然不得不提到,这种鉴相器也有一些问题。作为序列鉴相器,它对边沿丢失敏感。这样,它会把一个丢失的边沿误当作频率误差,并驱使环路去"校正"这个误差。另外,在零相位误差附近,鉴相特性的斜率实际上可能与图 16.13 所示的有点不太一样,因为在锁定点附近,鉴相器的 U、D 两个端输出的都是狭长的条形。再加上实际电路的速度都是有限的,非零的上升时间会使得长条形的面积不再和输入的时间(相位)差具有线性关系,因而鉴相器的特性会偏离上面所示的那种线性关系。

在有些系统里面,人们特意引入一个直流偏置来解决这个问题,这样锁住时鉴相器的输出是一个非零值。通过将平衡点偏置到离开鉴相器的中心,非线性问题可被大大地抑制。不巧的是,这个策略对于需要小误差的应用来说显然是不合适的,因为附加的偏置引入了一个静态的相位误差。

16.7 环路滤波器与电荷泵

到目前为止,我们用线性化模型研究了 PLL 的行为以及鉴相器的一些实现方法。下面要研究如何实现环路的剩余部分。会研究不同类型的环路滤波器,并将通过阐述一个实际

例子来说明典型的设计过程。

VCO 需要一个控制电压来产生一个需要的频率输出，为了能够从鉴相器的零输出(这样就有零相位误差)得到这个控制电压，环路里面需要一个积分器，并且，为了保证稳定，环路滤波器还要提供一个零点。图 16.14 给出了一个满足上述条件的经典锁相环结构。

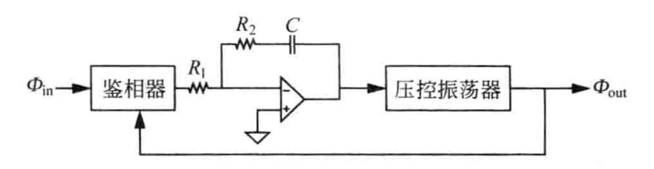


图 16.14　带有典型环路滤波器的 PLL

这个例子即使不推导一些方程也能很容易地得到它的特性。在频率非常低时，电容的阻抗控制了运放的反馈，因而环路滤波器的作用是积分器。但随着频率升高，容性电抗降低，最终等于串联电阻 R_2。在该频率以上，容性电抗与 R_2 相比越来越微不足道，并且增益最终趋向于 $-R_2/R_1$。

从另一个角度来看这个问题，就是一开始可以得到一个极点，然后可以得到一个时间常数为 R_2C 的零点。同时，还可以通过改变 R_1 来调节环路传输函数的幅度来符合要求，因此，可以控制整个运放电路来获得一个最终希望的环路传输函数。

在继续往前叙述之前，我们应该明白一个概念就是，PLL 并不一定需要如上所示的一个有源滤波器。最简单的情况，可以在鉴相器和 VCO 之间插入一个无源的 RC 网络，但是这样静态相位误差就不是零，而且环路带宽和静态相位误差是耦合的(成反比)。正因为以上限制，如此简单的环路滤波器只用在要求不高的场合。

图 16.14 所示的电路一般是用在分立电路中，但另一种电路(当然功能是一样的)用于许多其他的集成电路里。因为得到一个所期望的传输函数实际上并不需要一个完整的运放，所以采用一个比运放更为简化的元件可以大大缩小电路的复杂度和面积(如果不考虑功耗的话)。替代环路滤波器的一个比较流行的电路是“电荷泵”，它是和一个 RC 网络一前一后地工作着的。这样，鉴相器控制其中一个或者多个电流源，而 RC 网络则提供环路必要的动态范围。

图 16.15 说明了一个电荷泵是如何进行环路滤波操作的。这里，假设鉴相器提供一个“注入”或者“泄放”的数字信号。如果鉴相器发现 VCO 的输出滞后于输入参考信号，它就会激活上边那个电流源，把电荷灌入到电容上(注入)。如果 VCO 超前了，那么下边那个电流源就会被激活，把电容上的电荷抽取掉(泄放)。

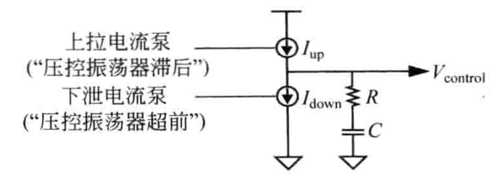


图 16.15　带有环路滤波器的基本电荷泵

如果没有电阻，则这是一个纯粹的积分器，而串联电阻却能够带来一个环路稳定所必需的零点，使得高频阻抗趋向一个非零值。

由于开关电流源用几个晶体管就很容易实现，因此如果要实现书上的运放那样的环路滤波特性，并且要综合考虑复杂度、面积、功耗等限制的话，电荷泵是一个很好的手段，而且它的控制特性又和很多现有的数字鉴相器刚好吻合（例如序列鉴相器），如同图 16.12 所示的那样。当那个鉴相器用于图 16.15 所示的电荷泵时，净泵电流为

$$I = I_{\text{pump}} \frac{\Delta\Phi}{2\pi} \tag{23}$$

其中 $I_{\text{pump}} = I_{\text{up}} = I_{\text{down}}$。这个电流乘以连接到电流源滤波网络的阻抗，就得到 VCO 的控制电压。

控制线上的抖动和高阶极点

即使电荷泵设计得很好，依然必须假设控制电压是有抖动的。使环路稳定的零点是以降低高频滤波性能为代价的。因而，在驱动 VCO 的控制线上存在着大量的高频成分。这些杂乱信号可能是来自于对电荷泵泄漏的补偿，也可能来自于乘法器型鉴相器的高阶混频项（例如倍频项），还可能来自于电荷泵或鉴相器的不对称性。VCO 频率受到的这种调制效应在很多应用中都是无法容忍的。许多这些分量都是周期性的，因而会产生固定的边带（杂散）。不巧的是，杂散很容易从注入到控制线的噪声中产生，包括电源、衬底中的噪声，甚至是外部场耦合到芯片中的噪声。典型的 RF VCO 的调谐灵敏度是几十或几百 MHz/V，所以即使几毫伏的噪声就能产生明显的频谱偏差。这导致在许多应用中，VCO 的频率调制是不可接受的。

实际电荷泵的设计可能比图 16.15 中的更难。其中的微妙之处可以从一个典型的设计中体会，如图 16.16。对这个 CMOS 电路的分析突出了一些与电荷泵设计相关的重要的设计考虑。晶体管 M1 到 M4 是由鉴相器高低电平指令控制的差分开关。这些命令的状态决定了是拉电流 I_{up}（source current）还是灌电流 I_{down}（sink current）流向输出结点 O_p。因此，I_{out} 等于 I_{up} 还是 I_{down} 取决于鉴相器的状态。

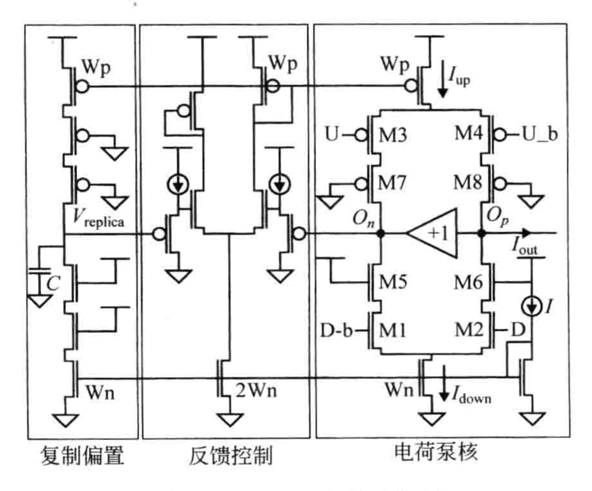


图 16.16　PLL 电荷泵例子

这些开关与晶体管 M5 到 M8 形成共源共栅连接，从而可以获得更高的输出阻抗，因为任何泄漏都会增大杂散的功率。为了理解为何会这样，我们来考虑锁相的情形。泄漏较低时，每个电荷泵周期里需要传输的静电荷很少。因此，控制线的抖动很小，从而对 VCO 的调制效应很小。然而，随着泄漏的增大，电荷泵必须补偿电荷流失量增大的影响，这意味着静态相位误差必然会增大。例如，如果泄漏是相位测量之间的控制电压差，那么相位误差会一直增大，直至上升沿脉冲存储的电荷比下降沿脉冲存储的电荷多到足可以补偿这个泄漏为止。共源共栅结构通过减小泄漏，可以帮助减小控制线抖动，因而减小了杂散的能量（以及静态误差）。由于在频率等于基准输入信号频率时产生的修正之间的电压降，控制线抖动也有与基准输入信号相等的基本周期。因而，杂散与载波间有偏移，偏移量等于基准信号频率。出现大的基准频率杂散，通常意味着电荷泵设计欠佳，如图 16.17 所示。可以清楚地看到，基准杂散位于距离 4.96GHz 载波 11MHz 的频率上。还有其他对应于控制线抖动的傅里叶（Fourier）分量的杂散信号（位于距离载波频率 11MHz 的整数倍的频率上）。

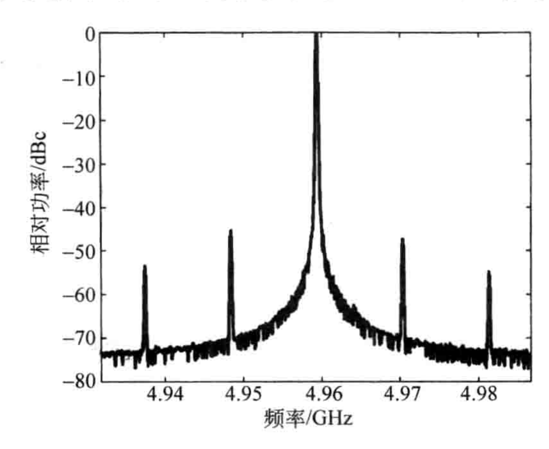


图 16.17　带有一些泄漏的电荷泵频率综合器的输出频谱

由于同样的原因，使上和下电流相等也是很重要的。如果其中一个比另一个大，就必然会出现补偿静态相位误差，以及对控制线抖动的伴生影响。为了缓解该问题，这里电荷泵设计采用了相对较大的器件（为了降低阈值不匹配），并使其工作于适度大的过驱动状态。另外，一个简单的单位增益缓冲器使得未使用的电荷泵输出与主输出具有相同的共模电压，从而消除因漏-源电压不等引起的系统失配。这种策略的一个补充是复制偏置环，其输出电压与未使用的电荷泵输出电压相比较。一个简单的运放驱动这两个电压到相等（补偿电容 C 是为了环路稳定性），并且因此保证了主核心的所有导通器件具有与复制偏置环相等的偏置电压。从而在随机失配的限制内，产生相等的上和下尾电流。

对这些细节的考虑可以很大程度上抑制基准杂散，从图 16.18 中可以明显看出。在这幅图中，杂散是看不到的，因而小于－70dBc 的噪底。基准杂散功率减小＞25 dB 代表了 300 倍的改善（在功率基础上）。

我们看到了环路滤波器在消除鉴相过程（如果你回忆一下，基本上是数字电路中的采样系统）产生的“牙齿”以及其他可能的耦合噪声时发挥的重要作用。我们要考虑如何设计最佳的环路滤波器。对于已知的环路带宽，高阶滤波器可以更大程度地衰减带外分量。然而，

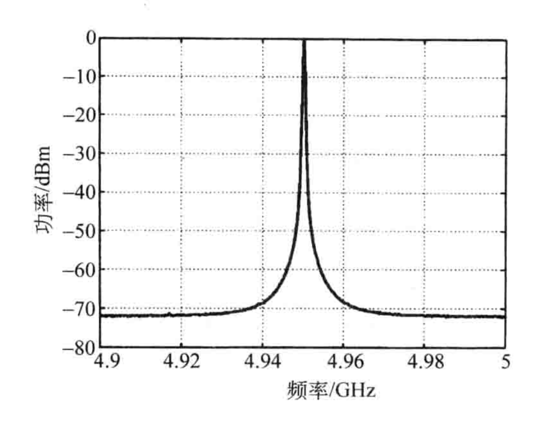


图 16.18　改进了的频率综合器的输出频谱

阶数越高，越难使环路稳定。因此，许多简单的频率综合器环路都是二阶的，但性能上很少有优势。设计高阶环路需要更加小心地应对不稳定的问题——也同样要求我们的数学能力比较强。根据收益递减法则，我们发现四阶的环路大约是最佳的，所以将关注三阶环路滤波器的设计，如图 16.19 所示。

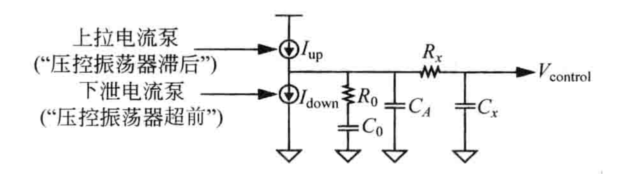


图 16.19　带有三阶环路滤波器的理想 PLL 电荷泵

回忆起 VCO(在原点)增加了另一个极点，我们看到选择一个三阶的环形滤波器造成了一个四阶环路。元件 C_A，R_X 和 C_X 在简单的二阶环路之外提供额外的滤波功能。在过去，并不存在一个简单的闭式设计方法，因此设计这样一个滤波器的过程要仔细看许多图，然后又放弃而回到一个二阶或三阶的环路。值得庆幸的是，最近这个状况开始改变了，我们能提供一个简单的，对大多数应用非常接近于最优化的"食谱"[⑩]。或者，许多常用的 CAD 工具(例如，National Semiconductor 的 PLL_LpFltr)可实现自动设计，可以得到比我们的"食谱"更好一些的环路滤波器。

步骤 1　定义一个相位冗余度。一旦这个值被选定，它就设置了电容值的限制。也就是，

$$(PM) \approx \arctan(\sqrt{b+1}) - \arctan\left(\frac{1}{\sqrt{b+1}}\right) \tag{24}$$

其中，

$$b = \frac{C_0}{C_A + C_X} \tag{25}$$

选择在目标值以上几度的相位裕度是明智的，因为这样可以对环路采样本性、未模拟的极

⑩　参考 H. R. Rategh and T. H. Lee, *Multi-GHz Frequency Synthesis and Division*, *Kluwer*, Dordrecht, 2001.

点，以及其他不稳定源起到缓冲作用。例如，假设定义的相位裕度目标是45°，如果因此设计50°，就会发现（如通过迭代）b应该大约是6.5。这是典型的设计值。

步骤2 选择一个跨越频率。比如说，这个频率是基于跟踪频带的指标规定。结合步骤1的结果，我们发现环路稳定零点的位置如下。

最大化环路带宽会使频率范围增大，在该频率范围内，想必优异的基准振荡器的相位噪声特性被赋予到输出上。不巧的是，环路是个采样数据系统，并且在离散时间鉴相器的固有相位延迟开始使相位裕度严重恶化之前，只能把交叉频率（crossover frequency）提高到大约是相位比较频率的十分之一。举个具体的例子，假设基准频率（以及相位比较频率）是2MHz。选择200kHz的交叉频率是保险的，因为这比基准频率低十倍频。如果将交叉频率定得比这个值高太多，那就太不靠谱了。

对于交叉频率，有

$$\omega_c \approx \frac{\sqrt{b+1}}{\tau_z} = \frac{\sqrt{b+1}}{R_0 C_0} \tag{26}$$

步骤3 计算C_0，即构筑零点的电容：

$$C_0 = \frac{I_P}{2\pi}\frac{K_0}{N}\frac{b}{\sqrt{b+1}}\frac{1}{\omega_c^2} \tag{27}$$

其中，I_P是电荷泵电流，N是分频比，K_0是VCO的增益常数，单位是rad/(s·V)。

步骤4 计算$R_0=\tau_z/C_0$。这就完成了环路滤波器主要部分的设计。

步骤5 在下面这个范围中选择$\tau_x=R_X C_X$：

$$0.01 < \tau_x/\tau_z < 0.1 \tag{28}$$

在这些宽松的限制内，有很多自由的选择。可以设计成算术平均值，或几何平均值，或是其他平均值。一般将τ_x选为τ_z的1/30～1/20。过大的时间常数会导致更好一些的滤波性能，但稳定性较差。由于环路常数不是恒定的，为谨慎起见，要留有一定的余量。

步骤6 完成余下的计算。回到步骤1，我们推出了电容比的约束条件。已经找到了其中一个电容，我们现在知道了C_A与C_X之和。可以在一个很宽的范围内任意选择它们各自的值，只要它们之和是正确的值。任意设定它们相等是一种常见的选择[11]。完成这些之后，就能确定它们的绝对值，随后又可以据此来确定R_X的值。

这就完成了环路滤波器的设计。

16.8 频率综合

带有高Q值谐振器的振荡器具有最好的频谱纯度，但相对频率的调谐范围不会超过10^{-4}左右。由于大部分收发机必须工作在分布较广的不同频率上，一种简单的应对调谐能力差的方法就是对每个频率采用单独的谐振器。显然，这种直接的方法只能用在所要求的频率数量不多时。

---

⑪ 滤波器中电阻产生的噪声会产生VCO的宽带调制，导致相位噪声。要使相位噪声最小化，就会在环路滤波器设计中引入另外的限制条件，我们将要考虑的全部就是这里所提供的“菜谱”式的步骤，这一事实使得情况足够复杂了。选择可以使整体实现较少地依赖于寄生的值是明智的。一般来说，采用与所要求时间常数一致的最大电容会有助于降低控制电压的宽带噪声调制。

实际上相反，所有的现代电子设备都采用了某种形式的频率综合，在其中，单个石英控制的振荡器与 PLL 以及其他数字元件相结合，提供多种输出频率。这些频率的信号可以追溯到高度稳定的基准信号。这样在理想情况下，就可以从一个振荡器获得宽广的工作频率范围以及良好的稳定性。

然而，在开始详细研究各种频率综合器之前，需要暂且研究一个严重影响架构选择的问题。在我们将要研究的频率综合器中都采用了一个分频器，恰当地模拟该分频器对环路稳定性的影响是很重要的。

16.8.1　分频器"延迟"

有时候，我们会在关于 PLL 频率综合器的文献中，在讨论环路稳定性时遇到"分频器延迟"这个术语。我们马上会看到这种现象的命名不太准确，但在有分频器的环路传输中确实存在稳定性的问题。

采用分频器一般意味着鉴相器本质是数字的[12]。因此，只有在离散的瞬间才有关于相差的信息。这就是说，环路是个数据采集系统。如果存在分频器，环路对相差的采样频次就可能会比 VCO 频率意味的要低。为了对 PLL 正确地建模，就需要恰当地描述它的采样特性。

为了深入理解，我们来考虑一个连续时间功能被周期地采样保持的过程，如图 16.20 所示。图中的采样保持操作会给该过程引入相位延迟。不需要用数学来解释，仅"观察"采样保持波形，并思考它与原连续波形的时间关系。你应该能够说服自己，两条曲线最佳的拟合是将原波形向右移动大约半个采样周期得到的。

更正式地讲，采样保持操作中的"保持"部分可以用一个脉冲响应是单位面积以及 T 秒持续时间的元件表征，如图 16.21 所示。这种元件的正式名称是零阶保持器(zero-order hold，ZOH)，其传输函数是[13]：

$$H(s)=\frac{1-e^{-sT}}{sT} \tag{29}$$

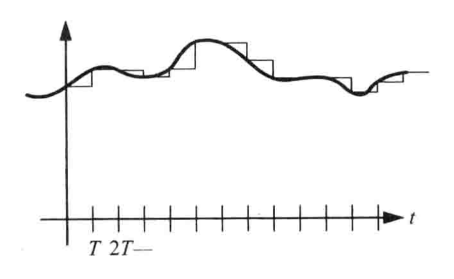




图 16.20　采样保持电路的操作

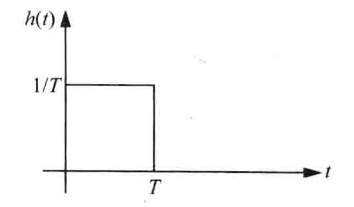


图 16.21　零阶保持的冲击响应

这个传输函数的幅度是：

$$|H(j\omega)|=\frac{\sin\omega(T/2)}{\omega(T/2)} \tag{30}$$

---

[12] 虽然有例外(例如，次谐波注入锁定振荡器)，但我们现在的讨论将仅限于常见的应用中。

[13] 如果你愿意快速地推导这个传输函数，要记住零阶保持器的脉冲响应与两个积分器之差相同(其中一个延迟时间)。这应该是个充分的提示了。

相位是：

$$\angle[H(\omega)] = -\omega(T/2) \tag{31}$$

因此延时是 $T/2$ 秒。根据对称性，可以知道大约以零秒为中心的矩形脉冲响应具有零相位，从而可以得到同样的结果。将响应向右平移 $T/2$ 秒，就得到如图 16.21 所示的图形。与我们如何计算无关，存在延迟的事实就是“分频器延迟”的原因。然而，幅度相对于频率不是常数，“延迟”这个词就不是完全正确的[14]。

现在，我们来把这个信息应用到具体的 PLL 例子中，该 PLL 的环路传输中含有分频器。为了表达相移，我们可以看到分频器可能会给环路带来的不利影响。随着分频比增加，采样周期 T 会增加（假设 VCO 输出频率固定）。这样，增加的负相移变得更加差，使相位裕度恶化。因此，为了避免这些影响，环路交叉必须控制在比 1/T 低的频率上。由于采样频率决定于分频器的输出，因此也取决于进行相位比较时的频率（而不是 VCO 的输出频率），高分频系数可能导致对环路带宽的严重约束，以及所有随之而来的设定速率和噪声性能的负面影响。因此，通常是将环路交叉频率选在大约是相位比较频率的 1/10。

最后，实际的分频器并不是无抖动的。也就是说，一个无噪声的输入并不产生无噪声的输出，因为必须对输出的恶化有所考虑。常用的一个经验法则是将频率综合器的输出噪声功率设计成比所需的低 2～3dB，来适应这个事实。这个经验法则只能用于不知道噪声幅度详细信息的场合。也就是说，额外的相位裕度对于许多实际设计都足够了。

16.8.2 静态模量频率综合器

在理解了环路传输中分频器带来的限制条件后，我们现在来研究各种拓扑结构的频率综合器。

图 16.22 所示为一个简单的 PLL 频率综合器，其中有一个基准振荡器和两个分频器。环路迫使 VCO 处于一个可以使 PLL 的输入信号频率相等的频率。因此，可以写出：

$$\frac{f_{\mathrm{ref}}}{N} = \frac{f_{\mathrm{out}}}{M} \tag{32}$$

所以有

$$f_{\mathrm{out}} = \frac{M}{N} f_{\mathrm{ref}} \tag{33}$$

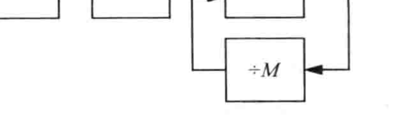


图 16.22　经典 PLL 频率综合器

这样，通过改变分频比（divide moduli）M 和 N，可以实现输入基准频率的任意有理倍数。输出的长期稳定性（即平均频率）与基准信号一样好，但短期稳定性（相位噪声）取决于净分频比，以及 PLL 中的 VCO 和环路的动态性能。在 PLL 环路带宽内，输出相位噪声将会是 M/N 乘以基准振荡器的相位噪声，因为相位相乘必然伴随着频率相乘。在 PLL 环路带宽外，反馈是无效的，因而输出相位噪声将等于 PLL 自身 VCO 的相位噪声。实际上，额外的噪声源（分频器和鉴相器）将会使综合后的相位噪声比理论最小值要大。因此，应将设计目标定为比要求的值更加严格 2dB。

注意到输出频率在步骤 f_{ref}/N 中可以增大，这个频率代表了 PLL 中鉴相器的工作速

[14] 幅度接近于 1，因为 $\omega T<1$。

率。稳定性的考虑以及抑制控制电压抖动的需要都要求必须采用比 f_{ref}/N 小的环路带宽。为了获得(假定)低噪声基准的最大利益,我们却希望 PLL 在尽可能宽的频带内跟随低噪声基准。另外,大的环路带宽会使分频比变化后更快地稳定下来。这些相互冲突的要求迫使人们研发替代结构的频率综合器。

有时会采用一种简单的改进,如图 16.23 所示。对于这个频率综合器,可以写出:

$$f_{out} = \frac{M}{NP} f_{ref} \tag{34}$$

显然,最小输出频率增量是 f_{ref}/NP,但环路在 f_{ref}/N 处比较相位,或者比之前的结构要快 P 倍。

但另一种改进是整数-N 频率综合器,如图 16.24 所示。在这种广泛采用的频率综合器中,分频器逻辑由两个计数器和双模预分频器(分频器)组成。一个计数器,称作信道间隔(吞咽)计数器,是可编程的,能够进行信道选择。另一个计数器,我们称其为帧计数器(也叫做程序计数器),通常是固定的并决定了包含如下操作的预分频器的总周期数:预分频器首先除以 $N+1$,直到信道间隔计数器溢出,然后除以 N,直到帧计数器溢出;预分频器分频比重置为 $N+1$,周期如此循环。

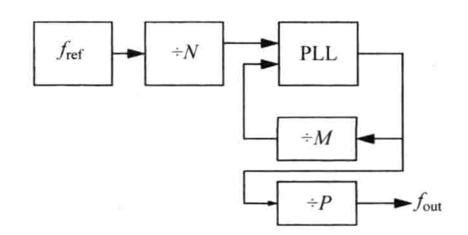


图 16.23　改进的 PLL 频率综合器

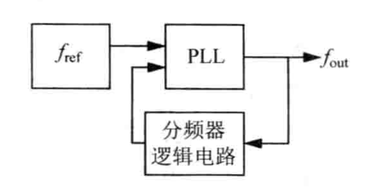


图 16.24　整数-N 频率综合器

令 S 为信道间隔计数器的最大值,F 是帧计数器的最大值。这样,在 S 周期内,预分频器将 VCO 的输出除以 $N+1$,在重复操作之前的 $F-S$ 周期内除以 N。因此,有效全局分频比 M 为

$$M = (N+1)S + (F-S)N = NF + S \tag{35}$$

因此输出频率增量等于基准频率。这个结构是实现图 16.22 所示基本原理框图的最受欢迎的结构。该结构是从其输出频率为基准频率的整数倍而得名的。

16.8.3　具有抖动分频比(dithered modulus)的频率综合器

在本书已经研究过的频率综合器中,期望的信道间隔直接限制了环路带宽。一种缓解该问题的方法是,在两个分频比之间抖动来产生比基准频率小的信道间隔,如图 16.25 所示。对其基本思想作一下说明,考虑到分频比在(例如)4 和 5 之间转换,50%的占空比等价于平均分频比为 4.5。变化任意一个分频比上所用的时间百分比,就能改变有效(平均)分频比,所以频率综合器输出可以以小于输入基准频率的频率步长增加。

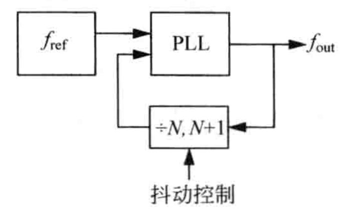


图 16.25　具有抖动分频比的频率综合器框图

有很多种在两个分频比之间的转换策略都可以产生相

同的平均分频比，当然，不是所有这些方法都是可取的，因为瞬时频率也是重要的。小数-N频率综合器中采用了最常用的策略，其中我们将 VCO 的输出在每个 K VCO 周期除以一个分频比（称作 $N+1$），在其他时间除以其他分频比（N）。因此，平均分频比为

$$N_{\mathrm{eff}} = (N+1)\left(\frac{1}{K}\right) + N\left(1-\frac{1}{K}\right) = N + \frac{1}{K} \tag{36}$$

所以，

$$f_{\mathrm{out}} = N_{\mathrm{eff}} f_{\mathrm{ref}} = \left(N + \frac{1}{K}\right) f_{\mathrm{ref}} \tag{37}$$

可以看到这种方法取决于 K，所以最小频率增量可以比基准频率小得多。然而，与目前研究过的其他频率综合器不同，鉴相器用频率比最小增量（实际上，鉴相器是由频率为 f_{ref}的信号驱动的）高得多的信号工作，因而提供了更期望的将综合器频率从 PLL 采样频率中去耦的方法。

为了更加详细地说明这种结构频率综合器的工作原理，我们来考虑用 10MHz 的基准输入产生 1.57542GHz 频率信号的问题。这种情况下，整数分频比 N 等于 157，而分数部分（$1/K$）等于 0.542。这样，（例如）对于每 1000 个 VCO 周期中的 542 个，我们希望除以 158（$=N+1$），而对于剩下的 458 个周期，我们希望除以 157（$= N$）。

在很多有可能实现这种性能的方法中，最常用的（但不一定是最佳的）一种方法是用每个周期的分频比的分数部分（这里是 0.542）来增加累加器的。每次当累加器溢出时（这里定义为等于或者超过 1），分频比就会被置为 $N + 1$。

溢出后剩下的量被保留，并且环路继续像以前一样工作。显然，分辨率是由累加器总累加量设定的，并且等于被累加器总累加量除的基准频率。对于本例 10MHz 的基准频率，一个 5 位 BCD 累加器就能使我们综合出的输出步长小到 100MHz。

分数-N 频率综合器还有一个值得提及的属性。由于环路是由在两个分频比之间周期切换的开关控制，必然会有对控制电压的周期性调制，因而也会对 VCO 的输出频率有周期性调制。因此，即使输出频率的平均值是正确的，也可能是不在一个瞬时基础上，输出频谱包含了边带。此外，边带的大小和位置都取决于具体的分频比和环路参数。

在实际的这种环路中，对这种调制的补偿一般是必需的。很多形式的补偿都是基于调制的决定性质——我们预先知道了控制线抖动的情况。因此，可以用变化的补偿控制电压来抵消掉不希望有的调制。实际上，这种技术（有时称为 API，即 analog phase interpolation）可以提供 20～40dB 的边带抑制。要实现更高水平的抑制就需要对 VCO 的控制特性非常了解，包括温度和电源电压效应，这些细节在不同的设计中都是大相径庭的[15]。

另一种抵消方法是采用两个相同的环路。在一定程度上，这两个环路匹配，两个频率综合区具有相同的抖动。其中一个提取抖动分量（如通过 DC 隔直电容），将其反向，然后再注入到第一个环路中。在稳态时，这种前馈校正就抵消了抖动分量。该方法的缺点是需要构造两个环路（使复杂度、面积和功耗都加倍），还要确保两个环路是完全相同的，并避免两个频率综合器产生不必要的相互作用。

也可以采用更加复杂的分频比切换策略，来消除控制电压周期抖动。例如，可以随机切

---

⑮ 参见 V. Mannassewitsch, Frequency Synthesizers, 3rd ed., Wiley, New York, 1987.

换,以提高噪底为代价来减小杂散频谱分量幅度。一种显著的改进是采用 delta-sigma 技术,来使噪声均匀分布[16]。如果要形成将噪声移至远离载波频率的频谱,随后的滤波器可以方便地将噪声滤除。环路本身会处理靠近载波的噪声,因此整个输出可以处理特别纯净的频谱。

16.8　合成频率综合器

另一种方法是将两个或多个频率综合器的输出合成起来。这种方法所提供的另一个自由度可以使折中的性能更为灵活,但这是以增加复杂度和功率耗散为代价的。

这种想法的最常用的表达方式是将固定频率源的输出与变化频率源的输出混合起来。失调频率综合器(offset synthesizer,如图 16.26 所示)就是这种特殊选择的一种实现形式。在这种结构中,环路并不伺服于输出与基准频率相等,因为附加的中间级混频使平衡点发生偏移。如果没有中间级混频,要注意平衡点将对应于(第一个并且唯一的)混频器之后的低通滤波器的零频率输出。这样,在补偿环路中,平衡点就对应于最后的低通滤波器的零频率输出。认识到了这点,就容易确定 f_{out}与 f_{ref}之间的关系了。

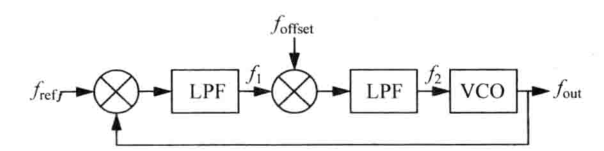


图 16.26　带有偏差频率输入的频率综合器环路

低通滤波器选择性消除了混频中产生的和频分量。因此,我们可以写出:

$$f_1 = f_{out} - f_{ref} \tag{38}$$

$$f_2 = f_1 - f_{offset} = f_{out} - f_{ref} - f_{offset} \tag{39}$$

设 f_2等于零,并且求解输出频率,得到:

$$f_{out} = f_{ref} + f_{offset} \tag{40}$$

这样,输出频率就是两个输入频率之和。

这种方法的一个重要的优点是输出频率不是基准频率的乘积。因此,相位噪声也同样不会相乘,这样就可以非常容易地产生一个低相位噪声的输出信号。相关的一个结果是对两个输入信号中任意一个的任何相位或频率调制都直接传递到输出信号,没有乘积因子的缩放。作为这些属性的结果,补偿频率综合器广泛用于 FM/PM 系统的发射机,特别是用于 GSM。

还有其他的方法来将两个频率的信号合成起来产生第三个信号。例如,可以用两个完整的 PLL,并用一个混频器将输出信号合并。为了选出和频分量而非差频分量(或者反之亦然),一般要采用一个滤波器。也可以采用一个单边带混频器(称作复混频器)来减小不希

---

[16] 关于这种结构的典型论文是由 Riley 等人发表的"Sigma-Delta Modulation in Fractional-N Frequency Synthesis",IEEE J. Solid-State Circuits, v. 28, May 1993, pp. 553-559. "delta-sigma"和"sigma-delta"这两个名词一般可以通用,但前一种命名方法是这个概念的发明者所使用的。

望有的分量的幅度，从而减轻滤波器的负担。然而，这种环路很少在IC电路中采用，因为很难避免两个PLL相互感应。一个常见的问题是两个环路通过衬底耦合或者经由放大器或其他电路的不完全的反向隔离（试图）锁住对方。这些问题都很难加以解决，所以目前很少采用这种双环路频率综合器。

16.8.5 直接数字频率综合

有些应用场合要求频率综合器具有较高速率的频率切换能力。比如在跳频扩频系统中，载波频率以伪随机形式变化[17]。传统的频率综合器很难提供满足要求的快速稳定性能，所以研发了其他方法。快速稳定频率综合器是开环系统，可以规避由反馈系统（例如PLL）的稳定条件带来的限制。

一种非常便捷的频率综合器设计方法是采用直接数字频率综合（direct digital synthesis，DDS）。图16.27所示为该结构频率综合器的基本框图。这个频率综合器包含一个累加器（ACC）、一个查表（带有整数输出寄存器）只读存储器（ROM）和一个数模转换器（DAC）。累加器接受频率指令信号（f_{inc}）作为输入，然后在每个时钟周期内将其输出增加这个量。因此输出线性增加，直到发生溢出以及周期重复。于是，输出Φ成锯齿形。认识到相位是频率的积分是非常有用的，所以累加器的输出类似于频率输入命令的积分。这样，锯齿形输出信号的频率就是时钟频率、累加器字长以及输入命令的函数。

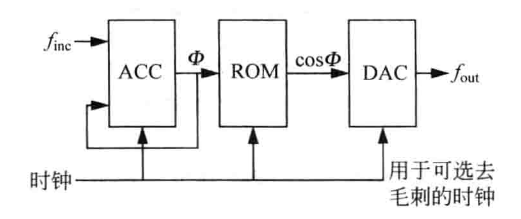


图16.27 直接数字频率综合器

然后，累加器的相位输出驱动ROM余弦查询表的地址线，该查询表将数字相位值转换成了数字幅度值[18]。最后，DAC将这些值转换为模拟输出。DAC之后一般要有一个滤波器，来将频谱纯度提高到可接受的水平。

简单地改变f_{inc}的值，频率就可以快速变化（有几个时钟周期），并且具有相位连续的行为特性。此外，对频率和相位都进行调制是通过在数字域内直接对f_{inc}和Φ调制来实现的。最后，甚至是幅度调制都可以通过采用乘法DAC（MDAC）来实现，其中模拟输出是模拟输出（这里，是指幅度调制）和ROM中的数字输出的乘积[19]。

这种频率综合器的主要问题是频谱纯度远不及之前讨论过的基于PLL的实现方法。DAC的位数是频谱纯度的一个限制（确切地讲，载波杂散比大约是6dB/b），而每个周期的ROM点数决定了谐波分量的位置（选择恰当的点数n，就能使第一个重要谐波出现在基波

[17] 在军事应用中，这种策略具有重要的反侦查和反干扰作用，这也是该策略的研发初衷。因为这样产生的频谱看起来非常像白噪声。

[18] 我们用一点点额外的逻辑，就可以简单地将所需ROM的数量降低75%，因为1/4的价值很容易重复使用，来重建整个周期。

[19] 在数字域也可以简单地实现幅度调制，就是在驱动DAC之前，将ROM的输出乘以所需幅度调制的数字表示。

$n-1$ 倍的地方)。由于时钟必须比最终产生的输出频率快很多,与基于 VCO/PLL 的频率综合器相比,这种频率综合器产生的信号频率要比现有技术的极限速率更小。通常,DDS 的输出是通过与基于 PLL 的频率综合器(或者是用作失调频率综合器的一个输入)的输出相混频来实现上变频的,从而在这两者之间达到折中。

16.9 一个设计实例

这里,我们将介绍一个设计实例,以便更好地理解前面的论述。假设想要设计一个输出频率为 1GHz 的频率综合器,从 25MHz 的基准振荡器开始。为方便起见,我们将采用现成的 IC(摩托罗拉的 MC12181,如图 16.28 所示),该 IC 包含了鉴相器、电荷泵、基准振荡器以及分频逻辑电路。我们所要做的就是连上一个晶体(为了基准振荡器)、一些辅助的无源元件和一个 VCO,然后再设计一个与之相匹配的环路滤波器。

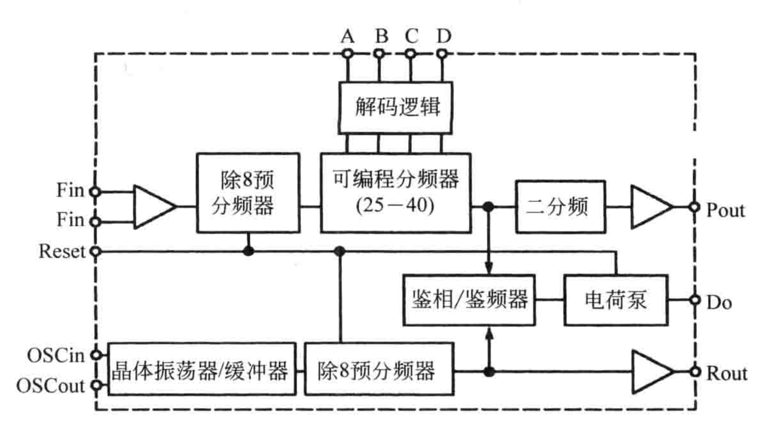


图 16.28 MC12181 频率综合器的简化原理框图(引自数据手册)

倍频因子可以设置为从 25～40 的 16 个整数值的任意一个,采用 4 个配置位。通过可编程因子将基准频率向上倍频,该因子是由连接该芯片两个针脚的外部晶体控制的。这里我们遇到了首个微妙之处:12181 内部的振荡器是采用传统的 Pierce 技术,就像共发射极放大器,虽然实际电路是更为复杂的。必须在晶体上外接一个约 50 kΩ 的电阻,来为振荡器建立偏置。此外,需要两个电容来提高额外的移相作用(超过晶体所能提供的),来满足振荡条件。如前面章节所述,Pierce 晶体工作在比其串联谐振频率稍高的频率,因而它代表了正常工作下的纯电感性阻抗。精确的振荡频率取决于晶体和这两个电容,虽然对于晶体的谐振频率要敏感得多(因为相对于电容,晶体的电抗-频率曲线要陡峭的多)。每个用于 Pierce 振荡器的晶体都被切割成仅在指定的容性负载频率才振荡。如果绝对频率精度很重要(因为一般都是这样),并且如果该容性负载不确定,那么这两个电容要是可调的,以便将振荡器调整到正确的频率。

高频晶体是非常薄的,因而很脆弱,并且很难加工。12181 的设计者们了解这点,因此指定 25MHz 作为频率上限,这个频率大概对应于制造商所提供的廉价的基波模式晶体的最大频率。不巧的是,25MHz 对于那些想用 8.33MHz 晶体来节约成本的制造商来说还是太高了,希望用户采用其三次谐波。而普通的振荡电路在基波和谐波频率上都可以满足振

荡条件，所以必须做出修改，来破坏基波模式频率下的环路条件，以免产生一些奇怪的、不希望有的结果。实现这种效果的最简单的方法是添加一个与电感串联的电感，从而在基波时产生一个串联电抗。该电抗产生一个到地的短路，使环路增益为零，于是阻止了在该频率产生振荡。为了使振荡器在三次谐波正常工作，需要在添加的电感两端并联一个电容。这里，我们假设已经有了一个 25MHz 的基波模式晶体，因而能够回避这个复杂的问题。

MC12181 立刻将晶体振荡器频率 8 分频，然后将分频后的信号送到鉴相器。因此，环路以 $f_{ref}/8$ 的速率进行相位比较，在此例中为 3.125MHz。为了避免考虑分频器延迟，将环路交叉频率设为比较频率的很小一部分。

现在假设采用图 16.19 所示的三阶环路滤波器来构造一个四阶 PLL。需要指定相位裕度和环路交叉频率。还需要知道将要连接到频率综合器芯片上的特定振荡器的 VCO 增益常数。鉴于本设计实例的目的，假设指定的相位裕度目标是 45°，环路将在 100kHz (628krps)处交叉，远低于 3.125MHz 的比较频率。最后，假设 VCO 恰好具有 100Mrps/V 的增益常数。有了这些数据，现在就按照前面阐述的步骤进行(在 16.7 节)。

步骤 1 我们指定相位裕度为 45°，所以保守地设计为 50°，b 约为 6.5，其中

$$b = \frac{C_0}{C_A + C_X} \tag{41}$$

步骤 2 选择了 628krps 作为交叉频率后，可以容易地找到与环路稳定零点对应的时间常数：

$$\omega_c \approx \frac{\sqrt{b+1}}{\tau_z} = \frac{\sqrt{b+1}}{R_0 C_0} \tag{42}$$

对于我们的数字，$\tau_z = R_0 C_0$ 计算出来大约为 4.4μs。

步骤 3 接下来计算 C_0：

$$C_0 = \frac{I_P}{2\pi} \frac{K_0}{N} \frac{b}{\sqrt{b+1}} \frac{1}{\omega_c^2} \tag{43}$$

对于 MC12181，额定电荷泵电流 I_P 为 2mA。对于我们的设计，N 是 320(要记得，在 4bit 可编程分频器之前，有一个内置的 8 分频预分频器)，K_0 是 100Mrps/V，所以 C_0 大约是 600pF。

步骤 4 我们有 $R_0 = \tau_z / C_0 = 7.3\text{k}\Omega$，完成环路滤波器的主要部分设计。

步骤 5 现在来设计附加的纹波滤波器。我们将任意地设置 $\tau_x = R_X C_X$ 为 τ_z 的 1/20，或者是 220ns。

步骤 6 设置了 C_0 之后，现在就知道了 C_A 与 C_X 之和，这里是 92pF。任意设两个电容相等，就得出每个电容是 46pF。最接近的标准值是 47pF，这与不太重要的计算值足够接近了。确定了这些电容之后，现在就能计算出 R_X 的值约为 4.7kΩ，这正好也是一个标准的 10% 的值。

很多 CAD 工具都能够用来设计三阶环路滤波器。例如 National Semiconductor 的 PLL_LpFltr 软件能够设计二阶和三阶环路滤波器，并且能给出 PLL 的环路传输增益和相位图。因为该软件采用了比这里概述的简单步骤更为复杂的算法，得到的环路滤波器元件值与我们计算出的值有些不同，如图 16.29 所示。

进行了环路滤波器所有的计算后，就完成了频率综合器的计算。

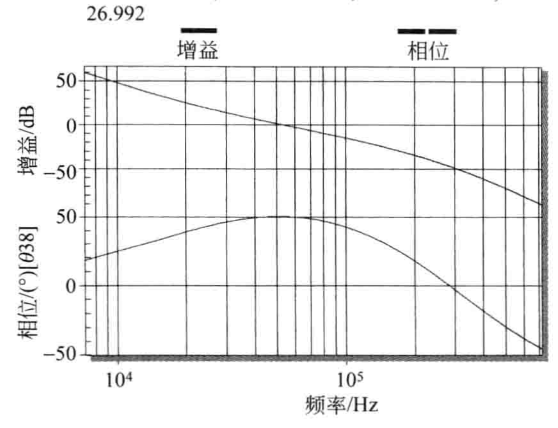


图 16.29　PLL_LpFltr 设计值和环路传输的波特图

实际问题的考虑

在上述设计实例中，假设 VCO 增益常数确实是常数。此外，隐含假设 VCO 的控制电压范围与 MC12181 的输出电压摆幅相匹配。最后，完全忽略了 VCO 控制端电荷泵的任何负载。在实际设计中，不是所有的假设都能满足，所以需要考虑实际 PLL 中的情况。

首先，非常数 VCO 增益意味着环路动态性能随控制电压变化而变化。为了避免在某些控制电压值上破坏环路的稳定性，环路滤波器必须保守设计，以适应 VCO 增益的最坏情况。这么做是非常繁琐的。许多设计者不进行这些繁琐的分析，而是采用频率综合器的宏模型，然后进行蒙特卡罗(Monte Carlo)分析来估计可能产生的相位裕度分布。值得庆幸的是，已经有合适的仿真工具了。例如，只要用对了宏模型，Spice 就能够实现这种方法。晶体管级模型对于实际仿真一般太过复杂了。

如果频率综合器芯片的输出电压范围不足以驱动 VCO，那么就需要对电荷泵的输出进行电平位移和放大。因为这个负载可能会降低稳定性，并引起杂散，所以也需要缓冲。这个附属电路必须具有足够宽的带宽，以免给环路传输带来较大的相位延迟。一个好的经验法则是确保增添的电路的第一个极点至少比控制电压范围内环路的最高交叉频率高十倍频程以上。同样，电平位移器的噪声必须保持较低，因为控制电压的任何噪声都能调制 VCO 的输出。在 PLL 带宽内，可以追踪这个噪声。然而在环路带宽外，这个控制电压噪声将会引起 VCO 输出频谱噪底的升高。

最后，VCO 控制端的电容要比 C_X 小，否则又会危及稳定性。

16.10　总结

这章讨论了很多频率综合器。对于稳定性的考虑使得环路交叉频率比相位比较频率低很多，而对于相位噪声的考虑有助于扩大环路带宽。因为在简单的结构中，输出频率增量与相位比较频率密切相关，很难以良好的增量来综合频率，同时给输出提供良好的相位噪声基准。小数-N 频率综合器消除了频率增量与相位比较率的关系，从而可以使用更宽的环路带宽。然而，相位噪声虽然是改善了，但控制电压的抖动会产生多种杂散分量。要抑制这些

杂散信号，可以通过消除抖动（因为在经典的小数-N 结构中，这是确定的），也可以通过采用随机分布或噪声型频谱。

16.11 附录 一个价廉的锁相环 PLL 设计实验指南书

设计一个微波频率 PLL 频率综合器可能需要相当昂贵的设计和测试硬件。使 PLL 工作在很低的频率，不花这么大的代价就能让学生了解 PLL 的工作原理。

我们研究的具体例子采用了市场上能买到的 PLL 芯片：4046 芯片。这是一种非常廉价（大约 0.25～1 美元）的 CMOS 器件，包含了两个鉴相器（一个异或门鉴相器和一个序列鉴相器）和一个 VCO。我们将研究怎么用每一个鉴相器和一些环路滤波器来设计一个 PLL。

虽然 4046 属于相对较慢的器件（最大振荡频率只有 1MHz），我们将要遵循的设计流程同样也适用于输出频率更高的 PLL，因此，下面的东西绝不仅仅是一个纯粹的学术练习。即使是今天，这个器件对很多应用来说仍然很有价值。因而它真的是一个非常便宜的教学用的 PLL 片上实验室。

16.11.1 4046 CMOS PLL 的特性

相位比较器（鉴相器）I

这块芯片包含有两个鉴相器。其中一个叫做“相位比较器 I”，是一个简单的异或门。回忆鉴相器那一节，异或门的增益常数是：

$$K_{\mathrm{D}} = \frac{V_{\mathrm{DD}}}{\pi}\ \mathrm{V/rad} \tag{44}$$

下面的设计例子里，都用 5V 作为电源，这样上式的数值就是：

$$K_{\mathrm{D}} = \frac{V_{\mathrm{DD}}}{\pi} \approx 1.59\mathrm{V/rad} \tag{45}$$

相位比较器（鉴相器）Ⅱ

芯片的另外一个相位比较器是一个序列鉴相器，工作在输入信号的上升沿，它有两个不同的工作状态，具体跟哪个输入超前有关。

如果输入信号边沿比 VCO 反馈信号边沿超前一个周期以内，那么当信号边沿到来时鉴相器的输出置高（就是 V_{DD}），当反馈信号边沿到来时变成高阻状态（一会儿我们就可以看到高阻状态很有用）。如果信号边沿落后于 VCO 的输出一个周期以内，那么当 VCO 边沿到来时鉴相器输出变低（接地），而输入信号边沿到来时变成高阻。以上就是这个鉴相器的工作状态。

高阻状态能够在锁住时减少控制线上面的电压波动量，这样锁频时不必要的相位和频率的调制效应就比其他类型的鉴相器要小的多。另外一点就是锁频时序列鉴相器得到的是一个零相位误差，这和用异或门得到的正交条件截然不同。

我们进行设计另一个需要考虑的信息是鉴相器的增益常数。不巧的是，由于高阻状态时它的输出电压与外部元件有关，而不是仅仅取决于相位误差，因此这个特殊的鉴相器没有一个准确的 K_{D} 的定义。解决这个问题的一个好方法是强制高阻时输出电压为 $V_{\mathrm{DD}}/2$（例如接一个简单的电阻分压器）。以此电路结构的改变，K_{D} 就可以算出来。

当相位误差小于一个周期时（输入信号超前于 VCO 信号），鉴相器的平均输出电压和相位误差成正比，其最小值是 $V_{DD}/2$（相位误差等于零时），最大值是 V_{DD}（相位误差等于 2π 时）。输出最小值实际上是由输出端接的电阻分压器决定的，而最大值则由电源电压决定。

类似地，对输入信号落后于 VCO 信号的情形，相位误差等于 0 时平均输出电压等于 $V_{DD}/2$，相位误差等于 2π 时平均输出电压等于零。这样，鉴相器的特性看上去就像图 16.32 所示的那样。求解带有近似边界条件的薛定谔方程（Schrödinger's equation）后，很容易就得出图中斜线的斜率是：



$$K_D = \frac{V_{DD}}{4\pi} \text{V/rad} \tag{46}$$

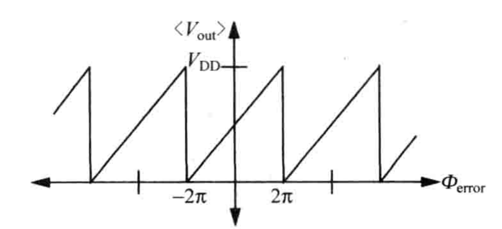


图 16.30 Ⅱ类相位比较器的特性

对于假设的 5V 电源电压 V_{DD}，鉴相器的增益大约为 0.40V/rad。

VCO 的特性

4046 里面的 VCO 让人们回想起在许多双极型 VCO 中用的发射极耦合多频振荡器。在电路中有一个外部的电容被一个电流源从两个不同的方向交替充电。当电容上的电压超过某个阈值时，一个简单的差分比较器就会变换电流源的极性。反馈的极性被选择为能够保持电路不断振荡。

VCO 的主要输出是一个方波，它是从差分比较器的一个输出端衍生出来的。在电容两端也可得到一个近似的三角波。如果需要一个正弦波的话，三角波是有用的，因为一个滤波网络或者非线性波形变换器都可以把三角波转换成一个类似正弦波的波形。

调整电容的充电电流就可以改变 VCO 振荡频率。通过选择外部的两个电阻，就可以分别调整 VCO 的中心频率和增益。电阻 R_2 用来设置 VCO 在没有输入情况下的充电电流（因而就设定了 VCO 频率），或者说在频率和控制电压的关系曲线上设定偏置点。另一个电阻 R_1 则通过设置一个共源极的跨导来调整 VCO 增益。

然而明显的是，在电路数据手册中并没有一个表来列出 VCO 频率和各种外部元件值的关系。下面这个准经验的公式（且是十分近似的）提供了这一很重要的信息[20]。在应用时要多加小心：

$$\omega_{osc} \approx \frac{2\left(\dfrac{V_C - 1}{R_1} + \dfrac{4}{R_2}\right)}{C} \tag{47}$$

从上面这个等式，对控制电压 V_C 求导就可以得到 VCO 的增益常数：

---

⑳ 这个公式是对一个用 5V 电源电压的个别器件测量的结果。这个公式的准确性因情况而异。尤其是用电阻值低于大约 50～100kΩ（VCO 的控制函数在更高的电流下变得十分非线性）。这里提醒一下书呆子们。

$$K_0 \approx \frac{2}{R_1 C} \text{ rad/(s} \cdot \text{V)} \tag{48}$$

附记

注意到鉴相器的增益是电源电压的函数，而且，VCO 频率也是 V_{DD} 的函数，因此，如果电源电压有波动，环路的动态反应就会变动。如果不想因为电源电压变动（包括噪声）而使环路行为受到影响，那么必须提供一个经过良好稳压和滤波的电源。为用户方便起见，4046 里面包括了一个 5.2V（±15%）的齐纳管，它可用于实现稳压的目的。

4046 还有一个简单的源跟随器来缓冲控制电压。当 PLL 被用在诸如 FM 解调器上这样的应用时，这个特点是很有帮助的。解调出来的信号就是 VCO 的控制电压，这样缓冲后的信号就可以直接驱动外部电路了。

最后，4046 内部还有一个"禁止"控制信号线，它能够关闭振荡器和源跟随器，来将芯片的功耗减小到 100μW 范围（如果输入信号是一个恒定的逻辑电平的话，功耗还能更小）。

16.11.2 设计实例

使用无源 RC 环路滤波器和鉴相器（PD）Ⅱ的二阶 PLL

我们知道有源滤波器可以提供优良的性能，尤其是稳态误差。然而，有一些应用场合全部使用无源的滤波器就可以了。在这种场合下如果用有源滤波器反而会白白增加面积和功耗。

假设要用鉴相器Ⅱ和简单的 RC 低通滤波器（没有环路稳定零点），来设计一个符合以下指标的电路：

交叉频率：1 krad/s

相位裕度：45°

中心频率：20kHz

解决方案：首先，注意到这个鉴相器的高阻抗特性需要使用一个电阻分压器，就像前面提到的那样。然后，为了让它能够去驱动一个任意的 *RC* 网络，最好是加一个缓冲器。这样就得到如图 16.31 所示的 PLL。

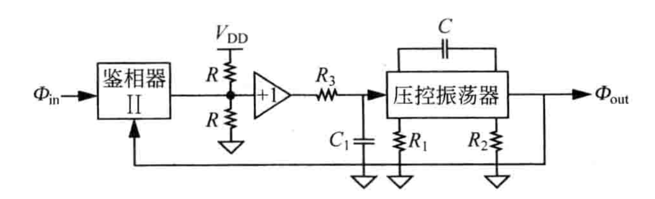


图 16.31 带有Ⅱ类相位比较器的 PLL

R 的具体数值并不十分关键，但是要足够大以保证鉴相器比较弱的驱动能力不会过载，大约 10 k 量级的数值就可以。环路传输函数可以写成：

$$-L(s) = K_D H_f(s) \frac{K_0}{s} = \frac{V_{DD}}{4\pi} \frac{1}{sR_3 C_1 + 1} \frac{K_0}{s} \tag{49}$$

由于没有环路稳定零点，相位裕度指标"要求"我们选择极点频率刚好等于所设计的交叉频率。作此选择后，必须选择 $R_1 C$ 来调整 VCO 增益。最后，选择 R_2 来满足中心频率的要求。

执行了以上这些步骤，同时注意为了使那个准经验的 VCO 等式成立，电阻值不能小于 50kΩ，就得到以下计算和估算过程：

(1)如前所述，相位裕度指标要求环路的滤波器的时间常数为 1ms。于是任意选取 $R_3=100\text{k}\Omega$，这样 $C_1=0.01\mu\text{F}$，刚好两个元件是标准数值。

(2)因为交叉频率必须等于 1krps，而且 R_3C_1 和鉴相器增益都已经知道，这样必须选择 K_0 使得交叉频率达到期望的频率：

$$|L(\text{j}\omega_c)|=K_D\frac{1}{\sqrt{2}}\frac{K_0}{10^3\,\text{rps}}=1\Rightarrow R_1C=0.582\text{ms} \tag{50}$$

任意选择电容等于一个标准值：0.001μF，这样电阻就必须是 582kΩ(不是标准值，但是接近 560kΩ 这个标准值)。作为参考，相应的 VCO 增益常数可以被计算出来，大约为：3.56krps/V。

(3) VCO 的电容如第(2)步选择好了以后，接下来选择 R_2 使得中心频率满足要求(这里定义控制电压等于 $V_{DD}/2$ 时 VCO 的频率为中心频率)。从准经验的 VCO 等式可以算出 R_2 大约要等于 67.3kΩ(最接近的标准值为 68kΩ)。考虑到器件特性的不一致性，如果 VCO 中心频率需要精确确定的话，R_2 最好选用在一定范围内以便调整。

以上就是设计的全过程。

有了上边的参数值，就可以计算 VCO 的可调范围、在这个可调范围内的稳态相位误差和锁定范围(这个以前没有讨论过)。锁定范围的定义是当输入端的频率变化时环路依然能锁住了，这个频率范围叫做锁定范围。

再次使用 VCO 公式来确定频率可调范围。根据上面的参数值，可以算出 VCO 的频率能够在中心频率上下 1kHz 的范围内变动，这个范围设定了 PLL 总体频率范围的上限。

由于采用了一个无源的环路滤波器，静态相位误差就不可能为零，因为一个零的相位误差需要鉴相器能产生一个非零的 VCO 控制电压[21]。如果假设 VCO 增益常数是一个不变量(即为常数)，那么就可以精确地计算出要使频率调节到上述步骤 1 计算出的那个范围上需要控制电压改变多少。如果相应的相位误差超过 $\pm 2\pi$ 范围，环路就不可能在整个 ±1kHz 的频率范围内保持锁定。

达到所要求的频率变化量所需的控制电压由 K_0 决定，这个控制电压又通过下式与鉴相器增益常数和相位误差相关联：

$$\Delta V_{\text{ctrl}}=\frac{\Delta\omega}{K_0}=K_D\Phi_{\text{error}} \tag{51}$$

用已确定的元件值，可以估算出在频率偏离中心频率 1kHz 时相位误差应该是 4.4 rad。实际测量表明，在频率下限(中心频率以下 1kHz)，相位误差是 4.3rad(理论计算结果看来是对的，译者注)。

然而，中心频率以上 1kHz 时，测量出来的相位误差实际上是约 5.9 rad。这个不小的差异来源于 VCO 的频率在控制电压较高时偏离线性较远。看起来要达到频率上限需要一个更高的控制电压，因此鉴相器也就需要一个更大的相位误差。可是 4.3 rad 和 5.9 rad 的相角都在鉴相器的线性范围内，因此应该是 VCO 的有限的调频范围而不是鉴相器的特性

[21] 这里，零控制电压指的是偏离 $V_{DD}/2$ 这个中值。

决定了这个具体例子中 PLL 的整体锁定范围。

使用无源 *RC* 环路滤波器和鉴相器(PD)I 的二阶 PLL

我们用异或门鉴相器代替序列鉴相器，重新做一下上述设计过程，是有意义的。由于异或门鉴相器的增益是 PD II 的四倍，因此 K_0 也应该缩小这个倍数来保证交叉频率不变。于是我们把 R_1 增大四倍来调整 K_0。为了保证中心频率 20kHz 不变，R_2 也要做调整(变小)。由于异或门没有高阻输出状态，电阻分压器和缓冲级都可以去掉。

所有的改变都完成了时，环路锁住时表现出来的行为和前面观察的都很类似。但是 VCO 的变化改变了它的可调范围，因此也改变了相应的相位误差：

$$\Delta V_{\text{ctrl}} = \frac{\Delta\omega}{K_0} = K_D \Phi_{\text{error}} \Rightarrow \Phi_{\text{error}} = \frac{\Delta\omega}{K_0 K_D} \tag{52}$$

因为 R_1 变大了，VCO 的可调范围降为原来的 1/4，而鉴相器增益和 VCO 增益的乘积保持不变。于是，异或门只是在序列鉴相器相位误差的 1/4 的范围内保持线性。因此，给定了交叉频率和衰减以后，异或门鉴相器使得环路具有更窄的锁定范围。

作为一个练习，读者可以自己进行数值计算来验证以上结论。(在这种情况下，好像 VCO 的可调范围仍然是限制因素，不过只是“仅仅”。)

异或门鉴相器最后要注意的几点，其中一点就是这种类型的鉴相器对于输入信号的占空比敏感。只有当两个输入信号都是 50% 占空比时，异或门理想的三角形特征才能保证。如果存在任何的非对称性，平均输出即使在相位误差处于极端位置时也不会达到电源电压或者地的水平。而序列鉴相器由于是一个边沿触发的器件因而没有这种占空比敏感性。

另外一个需要重申的很重要的问题就是，由于异或门操作相当于将两个正弦波相乘，因此它可能会锁在谐波上。方波所含有的丰富的谐波内容，使输入信号和 VCO 输出信号之间刚好有两个谐波频率一样的机会大大增加，这时就会锁住。如果不允许锁在谐波上，那么异或门鉴相器就会引起问题。

使用有源 *RC* 环路滤波器和鉴相器(PD)Ⅱ 的二阶 PLL

现在我们考虑把简单的无源 *RC* 环路滤波器替换成有源滤波器。同时要求滤波器在原点位置提供一个零点，以使稳态相位误差为零。假定交叉频率和相位裕度的指标与之前一样，但是要求环路在中心频率 ±10kHz 的范围内仍能够锁住。

为了满足相位裕度的要求，需要一个环路稳定的零点来抵消环路积分器带来的负相位贡献。这样，PLL 初看起来的样子如图 16.32 所示(VCO 的元件没有画出)。

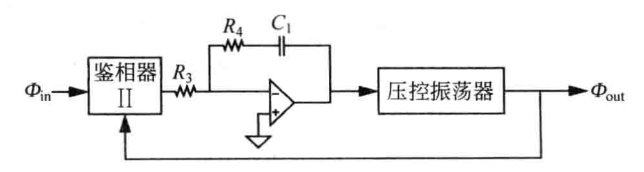


图 16.32　带有有源环路滤波器(有缺陷)的 PLL

为什么是“初看起来”？因为这个电路有一个小问题：如果输入信号比 VCO 超前，那么鉴相器的输出是正的，由于环路滤波器的反相特性会驱使 VCO 朝着频率更低的方向移动，加剧了相位误差，得到一个正反馈网络。为此，必须在控制线插入一个反相元件。

这个电路还有一个问题：运放的同相输入端是接地的，而鉴相器的输出最小时为地电平，这样环路滤波器就不会有积分输出了。解决最后这个(已知的)问题的方法是把运放的同相输入端接到 $V_{DD}/2$ 上，如图 16.33 所示。接下来就可以选取各个元件的数值了。

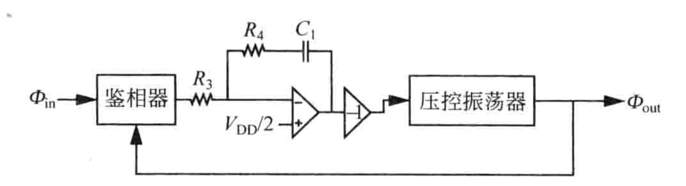


图 16.33　带有有源环路滤波器(已纠正)的 PLL

首先，环路的传输函数为

$$-L(s) = K_D H_f(s) \frac{K_0}{s} = \frac{V_{DD}}{4\pi} \frac{sR_4C_1 + 1}{sR_3C_1} \frac{K_0}{s} \tag{53}$$

从上式可以看到，在原点位置两个极点提供了总共$-180°$的相移，为了得到 45°的相位裕度，零点必须放在交叉频率的位置上。因此，R_4C_1 必须等于 1ms。其中一个元件值可以在一定范围内任取，例如取 $R_4 = 100\text{k}\Omega$，这样 $C_1 = 0.01\mu\text{F}$。

接下来，由于环路传输函数的幅度由 R_3 和 K_0 控制，因此如果只需要考虑交叉频率的指标要求，那么答案就不唯一。然而有另外一个考虑因素就是环路的锁定范围，这样 R_3 和 K_0 的选择是有限制的。更精确的讲，根据上述经验公式，只有控制电压在 1.2～5V 的范围内才对 VCO 输出有影响[22]。但这个电压中点是 3.1V，不是上面讲的 2.5V。因而若是仍然用 2.5V 作为电压中点的话，那么就无法得到一个对称的 20kHz 的区间，不过既然没有要求锁定范围非得要对称的，还是选择 2.5V 控制电压对应于 VCO 的中心频率。

根据上面的选择，VCO 频率下限是比频率上限范围小的，为了达到 10kHz 的指标，必须在控制电压为最小值(1.2V)时 VCO 频率至少改变 10kHz。1.2V 控制电压对应于偏移 1.3V，于是得到

$$K_0 > \frac{2\pi \cdot 10\text{kHz}}{1.3\text{V}} \approx 4.8 \times 10^4 \text{rps/V} \tag{54}$$

对于这个 VCO 增益常数，为了维持 20kHz 的中心频率，VCO 的各个元件取值如下：

$$C = 0.001\mu\text{F}, \quad R_1 = 42\text{k}\Omega, \quad R_2 = 130\text{k}\Omega$$

这里，最接近的标准值(10% 的容差)为：电阻 R_1 取 39kΩ，R_2 取 120kΩ。

最后，当其他元件都已经确定了以后，交叉频率要求运放的输入电阻必须满足：

$$R_3C_1 = \frac{K_DK_0}{\omega_c^2} \cdot \sqrt{2} \approx 27.7\text{ms} \tag{55}$$

因此，取 $R_2 = 2.8\text{M}\Omega$(最接近的标准值为 2.7MΩ)。这样，就完成了全部设计过程。

最后提一点，这个设计实例中，决定锁定范围的是 VCO 的可调范围，而不是鉴相器的特性。借助于环路滤波器的积分功能，零相位误差就能够给出稳态时 VCO 的控制电压。因此，鉴相器的特性与稳定时的锁定范围无关。

---

[22] 上述公式中，控制电压不允许是负的。

16.11.3 小结

在各种各样PLL可能的应用中，上述例子虽然是其中很小的一部分，但却是很有代表性。这里所涉及的频率低于1MHz，但基本原理对于微波频率都是有效的。将频率降低（3～4倍）可以让学生实践PLL设计，而不必头疼GHz频率PLL设计中的费用和夹具问题。

第17章 振荡器的相位噪声

17.1 引言

我们在第15章中已确认调谐振荡器比张弛(relaxation)振荡器的输出有更高的频谱纯度。一个直接的简单原因是一个高 Q 值的谐振器能很好地衰减中心频率以外的频率分量。这样,调谐振荡器可以更好地抑制失真,而且一个设计得好的调谐振荡器的输出波形能够更加接近于典型的正弦波。

除了可以抑制失真乘积项之外,谐振器还可以衰减由其他各种源产生的频谱分量,例如由于振荡器中的有限的 Q 值产生的热噪声,或由于在所有振荡器中都存在的有源器件产生的热噪声(由这些噪声源产生的噪声不仅会影响信号的幅度,而且会影响信号的相位,译者注)。在实际的振荡器中,限幅机制可以很好地削弱幅度的波动,因此相位噪声通常占主要地位,至少在离载波频率不远处的频率范围内是如此。因此,尽管可能设计出幅度噪声更为显著的振荡器,这里主要关注相位噪声。以后会指出这个有关相位噪声的理论经简单修改也可以解释幅度噪声。这样就可以精确地计算在频率偏离载波很远处的输出频谱。

不只是为了追求完整的分析,我们关注相位噪声的原因是使倒易混频(reciprocal mixing)的问题最小化。假若一个超外差接收器的本机振荡器完全是没有噪声的,那么两个相隔近的射频信号就是简单地一起下变频。然而,如果本振(LO)信号的频谱不是一个冲击函数,那么必须在实践上估计到一个非纯的LO频谱造成的后果。

在图17.1中,两个RF信号与LO信号外差而产生一对IF信号。所希望接收的RF信号比在一个邻近信道的信号要弱很多。假设(这也是典型的情况)前端滤波器没有足够的分辨度来进行信道滤波,下变频在将这两个RF信号平移至中频时,保留了原来的相对大小。因为LO频谱具有非零的带宽,下变频后的射频信号一样有带宽。LO频谱的这些带尾起了好像是在一个连续频谱上的许多寄生LO的作用。倒易混频是一些射频信号与这些不希望有的LO成分的混频。从图中可以明显看到倒易混频引起了不希望的信号在这个特定的情况下超过了希望接收的信号。为了使倒易混频的发生与严重程度降到最低,必须要减小LO的相位噪声。

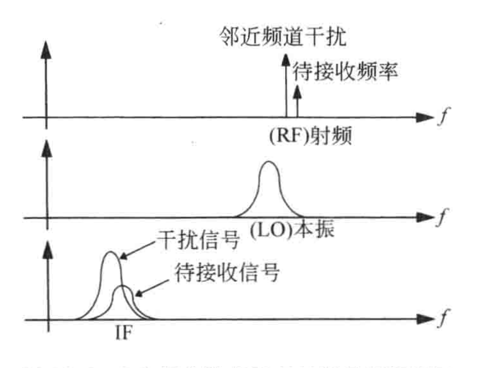


图17.1 LO相位噪声引起的倒易混频图示

振荡器的理论与实际的重要性已经激起了数不清的有关相位噪声的处理方法的发展。仅仅从就这个课题发表的文章的数目上就可以看出其重要性。同时,这些文章中的许多在相当基本的观点上互不相同的,可以这么说如此大量的互相矛盾的研究表明了许多这些处理方法是不合理的。下面这个事实又使得寻求一个合适理论的努力进一步复杂化,即电路

中的噪声在最终变为振荡器的相位噪声之前可能已经历了频率平移。这些频率平移经常被归结于在实际振荡器中存在的明显的非线性性。一些最简单的理论就简单地完全忽略了这些非线性性，并经常同时还忽略了(系统的)时变性。这种线性、时不变理论(linear time-invariant，LTI)能做到提供重要的定性设计内在知识的作用，但它们无疑在预测能力上是相当有限的。一个 LTI 理论的最主要的缺陷是频率平移必定是不允许的。这导致了这样一个问题的出现，即怎样来解释在实际振荡器中观察到的(近似)对称的边带是如何产生的。

尽管这个复杂性，也尽管存在着为了幅度的稳定所必需的非线性性，振荡器的噪声到相位的传输函数不管怎么说还是可以作为线性系统来处理。然而，一个有关频率平移过程的定量的理解要求抛弃时不变的原则，而这个原则在大多数有关相位噪声的理论中是隐含地假设的。除了提供一个在理论与测量结果之间的一致性解释，在本章介绍的时变相位噪声模型确定了一个重要的对称原理。这个原理可以被利用来抑制 $1/f$ 噪声的上变频至载波附近(close-in)的相位噪声。同时，它提供了使周期性稳态效应纳入理论框架的一个手段。这个周期性稳态效应在许多实际的振荡器中是十分明显的。同时这个对称原理也可以处理幅度到相位(AM-PM)的转换。这些内在的理解使得给予下面一个问题的重新解释成为可能：为什么某些拓扑结构如被广为推崇的考尔匹茨(Colpitts)振荡器有很好的性能。或许更为重要的是，这个理论能指导设计，对十分熟悉的振荡器提出新的优化方式，甚至能导致新的电路结构的发明。我们会通过仔细分析某些调谐 LC 振荡器电路例子来增强这种发展起来的理论考虑。本章以对一些实际模拟问题的一个简略考虑来结束。

我们首先需要重新提起这样一个议题，即如何来评价一个系统是否是线性的还是时不变的。这个问题很少在大多数系统的分析中出现，并且或许相当一部分工程师已经忘记了它们间的差别是什么。的确，我们发现甚至要考虑如何对“系统这个词意味着什么”给出明确的定义。然后确定一些非常通用的在关键参数间的折中选择。这些关键参数可以是功耗、振荡频率、谐振器的 Q 值和电路的噪声功率。然后，定量地研究如何在一个假想的理想振荡器中进行这种折中选择。在这个振荡器中，假设噪声至相位的传递函数是线性的，这样就允许通过冲击响应来进行表征。

尽管线性的假设是可以说得通的，我们会看到时不变性尽管在这个简单的情形下也不成立。就是说，振荡器是线性时变(linear time-varying，LTV)系统。这里系统是被定义为噪声到相位的转移特性。幸运的是，采用一个冲击响应来进行完全表征的过程只依赖于线性性，而不是时不变性。通过研究冲击响应，我们发现周期性的时变导致了器件噪声的频率平移以产生为实际的振荡器所显现的相位噪声谱。特别的是，可以看到 $1/f$ 噪声的上变频到近场噪声是依赖于对称性质，而这些对称性质是可以为设计者所控制的。此外，同样方式的处理很容易地包含了噪声发生器的周期性稳态。就如我们将会看到的，这个包容解释了为什么在一个振荡器里的有源器件处于 C 类操作可以是有利的。通过介绍一些有代表性的电路实例，可以强化我们对这个 LTV 模型的几个关键性的内在知识的了解。

17.2　一般性考虑

或许振荡器依然能保持与现实世界某种联系的最简单的抽象是一个带有损耗的谐振器与一个能量恢复元件的组合。后者恰好补偿了谐振槽路的损耗以维持一个等振幅的振荡。

为了简化问题的讨论，假设这个能量恢复器是不产生噪声的（见图 17.2）。槽路电阻因此是这个模型中的唯一有噪声的元件。

无噪声的能量恢复器

图 17.2　“具有完美效率”的 RLC 振荡器

为了得到一些有用的设计内在见解，首先计算一下存储在槽路中的信号能量：

$$E_{\text{sig}} = \frac{1}{2}CV_{\text{pk}}^2 \tag{1}$$

因此（载波）信号的均方电压值为

$$\overline{V_{\text{sig}}^2} = \frac{E_{\text{sig}}}{C} \tag{2}$$

这里已经假设了正弦波信号。

总的噪声电压均方值可以通过在 RLC 谐振器的噪声带宽内对电阻的热噪声谱密度进行积分来得到

$$\overline{V_{\text{n}}^2} = 4kTR\int_0^{\infty}\left|\frac{Z(f)}{R}\right|^2 \mathrm{d}f = 4kTR \cdot \frac{1}{4RC} = \frac{kT}{C} \tag{3}$$

结合方程(2)和方程(3)，得到噪声对载波信号的比值（为什么采用噪载比，而不是载噪比，这种上下颠倒的做法完全是出于习惯）：

$$\frac{N}{S} = \frac{\overline{V_{\text{n}}^2}}{\overline{V_{\text{sig}}^2}} = \frac{kT}{E_{\text{sig}}} \tag{4}$$

不言而喻，要取得最小的噪载比，必须使信号水平尽可能得大。

可以在公式中将功耗和谐振回路的 Q 值用显式表示出来。注意到 Q 值可以一般性地定义为正比于存储的能量除以消耗的能量：

$$Q = \frac{\omega_0 E_{\text{sig}}}{P_{\text{diss}}} \tag{5}$$

这样就可以写出：

$$\frac{N}{S} = \frac{\omega_0 kT}{QP_{\text{diss}}} \tag{6}$$

这个理想的模型振荡器消耗的功率就等于谐振槽路损耗的功率 P_{diss}。这里噪载比是反比于谐振回路的 Q 值与消耗功率的乘积，而且正比于振荡的频率。这些关系对许多实际的振荡器依然近似成立。这个观察也解释了，举个例子来说，为什么工程师对高 Q 值的执著追求。

通过将前面的这些讨论与有关实际振荡器的更多知识结合起来，其他的一些重要的设计准则就变得明显了。其中一个准则就是振荡器通常工作在以下两个区域中的一个。这两个区域是依振荡输出幅度对偏置电流的不同依赖关系而区分的（见图 17.3）。这可以写作

$$V_{\text{sig}} = I_{\text{BIAS}}R \tag{7}$$

其中 R 是一个具有电阻量纲的比例常数。这个常数又正比于等效的谐振回路并联电阻，因此

$$V_{\text{sig}} \propto I_{\text{BIAS}}R_{\text{tank}} \tag{8}$$

这意味着载波功率可以表示成

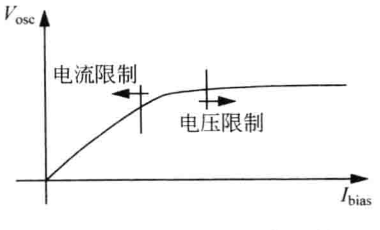


图 17.3　振荡器的工作区域

$$P_{sig} \propto I_{BIAS}^2 R_{tank} \tag{9}$$

噪声电压的均方值已经用谐振回路的电容来计算了：

$$\overline{V_n^2} = \frac{kT}{C} \tag{10}$$

但也可以用回路电感表示为

$$\overline{V_n^2} = \frac{kT}{C} = \frac{kT}{1/\omega_0^2 L} = kT\omega_0^2 L \tag{11}$$

因此，在电流限制的工作区域中的噪声对载波信号比的一个等价表达式是

$$\frac{N}{C} \propto \frac{kT\omega_0^2 L}{I_{BIAS}^2 R_{tank}} \tag{12}$$

假设工作在一个固定的电源电压下，对功耗的约束意味着偏置电流有一个上限。这样，在余下的可以自由选择的参数中只有谐振回路的电感和电阻能够被在实际中改变以取得 N/C 的最小值。亦即，这样一个振荡器的优化归结为最小化 L/R_{tank}。许多处理方法通过使回路电感最大化以实现优化。然而，我们看到了一个更为对头的目标是最小化 L/R_{tank}①。因为一般来说，电阻本身就是电感的函数。由此可知，要确定然后再得到这个最小值是不容易做到的。另一个需要考虑的因素是当电感值低于某一个最小值时，振荡会停止。因此，这里提出的优化策略是假定振荡一定会发生，且振荡器工作在输出幅度正比于偏置电流的区域。

17.3 更详细的讨论：相位噪声

为了深化前面分析中得到的定性的理解，我们现在来决定理想振荡器的实际输出频谱。

假定图 17.2 的输出是所示的谐振回路两端的电压。我们已经设定，电路中唯一的噪声源是回路电导的白热噪声。该噪声可以表示为并联在回路两端的电流源，其谱密度的均方值是

$$\frac{\overline{i_n^2}}{\Delta f} = 4kTG \tag{13}$$

这个电流噪声乘上电流源两端的有效阻抗就变成了电压噪声。然而在计算这个阻抗时，必须考虑到能量恢复元件也贡献一个平均的有效负电阻。这个负电阻正好抵消了回路的正电阻。这样，净效果是为噪声电流源看到的有效阻抗是一个理想的无损 LC 网络的阻抗。

对于与中心频率 ω_0 的一个很小的频率偏移 $\Delta\omega$，LC 谐振回路的阻抗可以近似为

$$Z(\omega_0 + \Delta\omega) \approx -\mathrm{j} \cdot \frac{\omega_0 L}{2(\Delta\omega/\omega_0)} \tag{14}$$

可以通过结合一个对未加负载的谐振回路的 Q 值表达式，把这个阻抗写成一种更有用的形式：

$$Q = \frac{R}{\omega_0 L} = \frac{1}{\omega_0 GL} \tag{15}$$

① D. Ham and A. Hajimiri, "Concepts and Methods in Optimization of Integrated LCVCOs"(集成 LC 压控振荡器优化时的设计的概念与方法), *IEEE J. Solid Circuits*, June 2001.

将式(15)中的 L 代入式(14)可以得到

$$|Z(\omega_0+\Delta\omega)| \approx \frac{1}{G} \cdot \frac{\omega_0}{2Q\Delta\omega} \tag{16}$$

这样就将对电感的显式依赖关系用对 Q 和 G 的依赖关系取代了。

下一步,将噪声电流均方值的谱密度乘以谐振回路阻抗幅度的平方来得到噪声电压均方值的谱密度:

$$\frac{\overline{v_n^2}}{\Delta f} = \frac{\overline{i_n^2}}{\Delta f} \cdot |Z|^2 = 4kTR\left(\frac{\omega_0}{2Q\Delta\omega}\right)^2 \tag{17}$$

由于谐振回路的滤波作用,输出噪声的功率谱密度与频率有关,且与偏差频率的平方成反比关系。这个 $1/f^2$ 行为简单地反映了这样一个事实,即一个 RLC 谐振回路的电压频率响应以 $1/f$ 的方式从中心频率两侧滚降,而功率是正比于电压的平方。同时也注意到,当其他参数不变时,回路 Q 值的增加降低了噪声频谱密度。这又一次强调了增加回路 Q 值的意义。

在我们理想化的 LC 模型中,热噪声在幅度和相位上都会引起扰动。式(17)将这两项效应都包括在内了。由热动力学的均分理论可知,在没有限幅的条件下,噪声能量体现在幅度和相位上是相等的。然而,在所有实际的振荡器中存在有幅度限制机制,因此幅度噪声被消除,剩下的噪声能量只有方程(17)给出的一半。

此外,与噪声的绝对数值相比,我们常常更加关心的是噪声与载波信号的相对大小。所以通常将均方噪声电压密度值用均方载波信号电压进行归一化处理,得到它们比值的分贝数,这就解释了先前采用的“上下颠倒”的比值。该归一化过程产生下面的相位噪声方程:

$$L\{\Delta\omega\} = 10\log\left[\frac{2kT}{P_{\text{sig}}}\left(\frac{\omega_0}{2Q\Delta\omega}\right)^2\right] \tag{18}$$

噪声的单位这样就正比于谱密度的对数(log)。具体来讲,相位噪声的单位通常表示为相对于某个载波频率 ω_0 偏移 $\Delta\omega$ 时,单位频率的噪载比分贝数,也就是 dBc/Hz。比如说,一个 2GHz 振荡器的相位噪声可以被说成是“在 100kHz 偏差下 −110dBc/Hz”。一味追求完美的人可能会抱怨“每赫兹”实际上是指 log 里面的变量,而不是 log 本身。因而,将测量频带加倍并不会使分贝数翻倍。但是就像“dBc/Hz”不十分严密一样,人们还是这么用的。

式(18)说明,在给定偏差频率处的相位噪声随着载波功率和 Q 值的增加而减小,正如先前预测的。这种依赖关系是符合逻辑的。增加信号的功率就减小了噪载比,因为热噪声本身是固定的。增加 Q 值可以成平方关系地减小这一比值,因为谐振回路的阻抗呈 $1/Q\Delta\omega$ 关系下降。

因为在得出以上的结论过程中作了许多简化了假设,发现为式(18)所预测的频谱与实际测量的典型结果有显著的区别,就不会令人惊讶。例如,尽管实际的频谱中的确存在着一个区间,其中测量到的谱密度正比于 $1/(\Delta\omega^2)$,但其大小一般要比式(18)给出的大许多。这是因为振荡器中除了谐振回路损耗以外,还有其他一些重要的噪声源。举例说,任何一个能量恢复电路的物理实现都会带来噪声。此外,测量到的频谱最终会在大的频率偏差下变平,而不是随着 $\Delta\omega$ 增加呈平方率下降。这个噪声下限可能是由于位于谐振回路和外界之间的有源器件(例如缓冲器)带来的,也可能反映了测试仪器本身的限制。即使直接从谐振回路两端得到输出,任何与电感或电容串连的电阻都会导致一个下限。这个下限限制了谐振回

路在大的频率偏差时能提供的滤波效益，并最终产生了一个噪底。最后我们要提到，对于小的频率偏移，总会有一个$1/|\Delta\omega|^3$依赖关系的区域。

对式(18)的修正提供了如下一个计入以上这些不一致处的公式：

$$L\{\Delta\omega\} = 10\log\left\{\frac{2FkT}{P_{\text{sig}}}\cdot\left[1+\left(\frac{\omega_0}{2Q\Delta\omega}\right)^2\right]\cdot\left(1+\frac{\Delta\omega_{1/f^3}}{|\Delta\omega|}\right)\right\} \tag{19}$$

这些由 Leeson 提出的修正包括：引入一个因子F来计入$1/(\Delta\omega)^2$区域增加的噪声，一个附加项1(中括号里)来引入噪声下限，一个乘数项(最后的圆括号中)来在足够小的偏移频率下提供$1/|\Delta\omega|^3$项[②]。有了这些修正，相位噪声谱如图 17.4 所示。

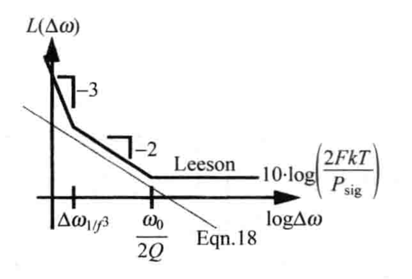


图 17.4　相位噪声：Leeson 模型与式(18)的比较

Leeson 模型是非常可贵的，因为它能够让人直观地理解振荡器。然而，应当注意因子F是一个经验拟合参数，因此必须从测量结果中得到。这样，相位噪声公式的预测性就变差了。此外，这个模型通常取区分$1/(\Delta\omega)^2$和$1/|\Delta\omega|^3$区域边界的频率偏移$\Delta\omega_{1/f^3}$为器件噪声的$1/f$的拐角频率，测量数据却往往不支持这样一个等式。因此，必须将$\Delta\omega_{1/f^3}$当作一个拟合参数。再有，当存在着不止一个噪声源时，应该采用什么样的拐角频率也是不清楚的。这些噪声源都呈现出$1/f$特性(但一般具有不同的拐角频率)。最后要提到的是，噪声变平的频率点也不总是等于谐振回路带宽的一半，$\omega_0/2Q$。

理想的振荡器模型和 Leeson 模型，二者都指出了提高Q值和信号功率是减小相位噪声的途径。Leeson 模型还引入了因子F，但并不清楚它依赖于什么机理。因此要找到减小F的方法也是十分困难的。同样的问题对于$\Delta\omega_{1/f^3}$也存在。最后需要指出的是，盲目应用以上这些模型，不断地导致某些设计者试图用有源电路来提高Q值，但这是一种误导。不幸的是，通过这种方法(即采用有源电路)来提高Q值必然伴随着F的增加，因为有源器件本身也产生噪声。这样，预期的相位噪声的改善并不会实现。再有，由于缺乏F的解析表达式，以上的结论看起来可能不明显。我们还是不断看到基于有源Q值增强方法的各种失败的振荡器设计。

表达式(18)或式(19)都不能定量地预测相位噪声这个事实，表明了在推导过程中用到的假设至少有一些是不正确的，尽管它们看起来合理。为了形成一个克服了以上例子中不足之处的理论，需要重新评估，甚至修改这些假设。

17.4　线性度与时变在相位噪声中的作用

在前面的推导中，都假定了线性和时不变性。让我们来逐个考虑这些假设。

在这一节，时不变性扩充到噪声源本身。对于噪声源，隐含假定了它处于稳态，意味着假定用来刻画噪声的描述性的度量(例如谱密度)也是时不变的。

---

② D. B. Leeson, "A simple model of feedback oscillator noise spectrum"(一个简单的有关带反馈振荡器的噪声谱的模型), *Proc. IEEE*, v. 54, February 1966, pp. 329-330.

很明显，非线性是所有实际的振荡器的一个基本性质，因为它的存在对于限幅是必要的。因此，试图将某些振荡器的特性用非线性行为来解释看起来是完全有道理的。这些特性中的一个是当一个单频率正弦波扰动注入到振荡器后，会引起两个对称地分布在载波两边的等幅边带③。因为 LTI 系统不能实现频率变换，而非线性系统可以，非线性混频经常被用来解释相位噪声的产生。就如我们很快会看到的，幅度控制的非线性性肯定能影响相位噪声——但是以非直接的方式，这个影响是通过控制输出波形的形状细节来施加的。

一个重要的内在理解是扰动就是叠加在主振荡上的扰动。对任何值得使用或分析的振荡器，它们在幅度上总是比载波信号小很多。因此，如果一定量的注入噪声产生了一定的相位扰动，我们应当期望将注入的噪声加倍，这个扰动也会加倍。就噪声-相位转移函数而言，线性因此看起来是合理的(并且是实验可以测到的)。因此这特别重要，当我们评估线性时，必须要显式地确定输入-输出变量。同样重要的是认识到这个线性假设并不等价于否认有源器件的非线性行为。因为这是围绕着稳态解的线性化，因此这样做已经将器件的非线性考虑在内。这种做法与放大器的分析是完全类似的，在那里小信号增益是在采用大信号(非线性)方程得到的在某一个偏置下的解附近定义的。因此这里与以前关于非线性幅度控制的知识没有冲突。任何看上去的矛盾是因为“系统”这个词实际上没有被确切地定义这个事实。许多人将系统认为是指电路元件及它们之间的连线的组合，但是一个更为有用的定义是基于被选择的特定的输入-输出变量。采用这个定义，一个给定的电路在某些变量之间可能具有非线性关系，而在其他一些变量间可能是线性的。时不变性也不是整个电路的本征性质，它类似地也是依赖于选定的变量。

现在，只剩下去重新仔细考察时不变这个假设。在前面的推导中，将时不变扩展到噪声源本身，即指表征噪声的测量(比如，谱密度)是时不变的(稳态)。与线性不同，时不变的假设是不那么明显可辩解的。事实上，来演示振荡器从根本上说是时变系统倒是非常简单。认识到这个事实是形成一个有关相位噪声的更精确理论的关键所在④。

为了测试时不变性是否成立，专门考虑一个冲击电流是如何影响一个最简单的谐振系统，即一个无损耗的 LC 谐振回路(图 17.5)。假设系统已经在以某个恒定幅度振荡了无穷长的时间，然后考虑这个系统是如何对在两个不同时间注入的一个冲击脉冲产生响应的，这个情况见图 17.6。

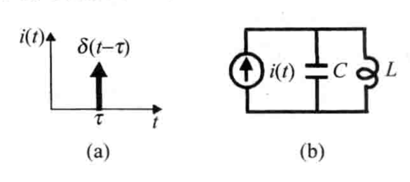
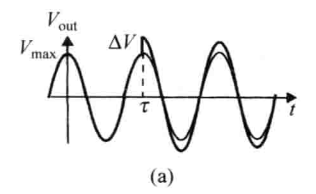


图 17.5 为电流脉冲所激励的 LC 振荡器

如果冲击脉冲碰巧在电压幅度最大时(图 17.6(a))注入，其突然增加的幅度为 $\Delta V=\Delta Q/C$，但因为对冲击信号的响应正好与原先存在的振荡完全同相位叠加，波形的过零点时刻并不会改变。这样，即使是明显已经改变了系统的能量，这个幅度上的变化是不会伴随有相位上的变化。在另一方面，在一些其他时刻注入的冲击信号，通常会既影响振荡的幅度，又会影响过零时刻，这在图 17.6(b)可以见到。

③ B. Razavi, "A study of phase noise in CMOS oscillators"(对 CMOS 振荡器的相位噪声的研究), *IEEE J. Solid-State Circuits*, v. 31, no. 2, March 1996.

④ A. Hajimiri and T. Lee, "A general theory of phase noise in electrical oscillators"(在电学振荡器中的相位噪声的一个通用理论), *IEEE J. Solid-State Circuits*, v. 33, no. 2, February 1998, pp. 179-194.



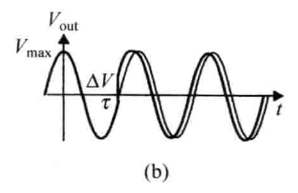


图 17.6　*LC* 谐振回路对冲击激励的响应

将过零时刻解释成一种对相位的度量，我们看到由一个给定的注入的冲击造成的相位扰动的量取决于这个冲击是什么时候发生的，因此时不变性就不再成立了。振荡器因此就是一个线性，同时是(周期性的)时变(LTV)系统。特别重要的注意到理论是有可能使系统的能量不发生变化(这就像在 17.6 右图上的恒定的谐振回路振幅所反映出来的)，这种情况发生在冲击是在靠近过零点时刻注入，这时为冲击注入所做的功为零。举个例来说，当谐振回路电压为负时，一个注入的正冲击脉冲从振荡器中汲取能量，而当谐振回路电压为正时，同样的这个冲击给振荡器输送能量。恰好在零点跨越的时刻之前，可以找到一个瞬间，此时注入的冲击完全不做功。因此，振荡的振幅不会被改变，但是过零点则发生了偏移。

因为(噪声至相位的转换的)线性保持为一个好的假设，这个冲击响应依然完全表征了那个系统——尽管存在着时变性。相对于一个 LTI 冲击响应的仅有的差别是这里冲击响应是两个自变量的函数：一个自变量是观察的时间 t，另一个是激发的时间 τ。因为一个冲击输入产生了一个在相位上的阶跃变化，所以冲击响应可以写作：

$$h_{\phi}(t,\tau)=\frac{\Gamma(\omega_0\tau)}{q_{max}}u(t-\tau) \tag{20}$$

这里，$u(t)$是单位阶跃函数。再除以电容两端的最大的电荷量 q_{max} 以使得函数 $\Gamma(x)$独立于信号的幅度。这个归一化过程使得我们能方便地客观比较各种不同的振荡器。注意，这个 $\Gamma(x)$被称为冲击灵敏度函数(impulse sensitivity function, ISF)，是个无量纲的与频率和幅度无关的周期函数，其周期为 2π。从命名可以看出，它包含了系统在相位 $\omega_0\tau$ 处对冲击注入的响应。在我们的 *LC* 振荡器的例子中，$\Gamma(x)$在振荡器的过零点附近存在最大值，在振荡波形的峰值处为零。一般来说，通过模拟来决定 $\Gamma(x)$是最可行的(也是最准确的)，但是也有适用于特殊场合的解析方法(要作某些近似)⑤。不管怎么说，来培养一个对 ISF 的典型形状的感觉，考虑两个有代表性的例子：图 17.7 所示的 *LC* 振荡器与环形振荡器。

一旦冲击响应被确定下来(无论以任何形式)，可以用叠加积分的方法算出因为一个任意的噪声信号引起的过量相位。这里，这样的计算是合理的，因为叠加这个特性是与线性有关的，而不是与时不变性相关联的：

$$\phi(t)=\int_{-\infty}^{\infty}h_{\phi}(t,\tau)i(\tau)\mathrm{d}\tau=\frac{1}{q_{max}}\int_{-\infty}^{t}\Gamma(\omega_0\tau)i(\tau)\mathrm{d}\tau \tag{21}$$

图 17.8 所示的等效方框图帮助我们将这个计算过程图示化，这个图对于电信工程师来说是颇为熟悉的，他们会认得出这是一个超外差系统(我们很快还要回到这个观点上来)。

---

⑤　F. X. Kaertner, "Determination of the correlation spectrum of oscillators with low noise"(决定振荡器的频谱与低噪声的关联), *IEEE Trans. Microwave Theory and Tech.*, v. 37, no. 1, January 1989. 同时参看 A. Hajimiri and T. Lee, *The Design of Low-Noise Oscillators*(低噪声振荡器的设计), Kluwer, Dordrecht, 1999.

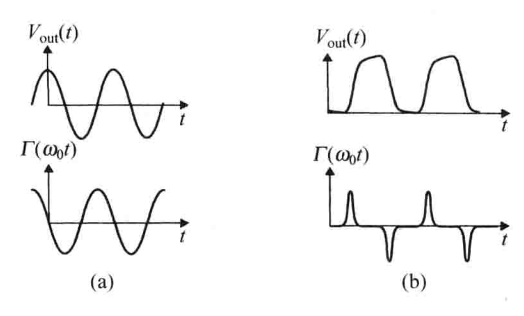


图 17.7　冲击灵敏度函数(ISF)实例

(a) LC 振荡器；(b) 环形振荡器

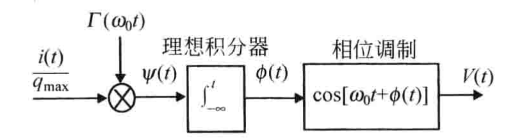


图 17.8　为式(21)部分描述的过程的等效方框图

为了把这个叠加积分写成一个更加实用的有用形式，注意 ISF 是周期函数因而可以用傅里叶级数展开：

$$\Gamma(\omega_0 \tau) = \frac{c_0}{2} + \sum_{n=1}^{\infty} c_n \cos(n\omega_0 \tau + \theta_n) \tag{22}$$

这里系数 c_n 是实数，θ_n 是 ISF 第 n 阶谐波的相位(在后面的讨论中，将忽略 θ_n，因为假设噪声成分互相间是不相关的，因此它们的相对相位是无关的)。采用这个展开的价值在于——就像许多与实际物理现象相关的函数——这些级数的收敛一般是很快的，所以通常可以只用级数的前几项就可以很好地近似整个函数。

将傅里叶展开代入式(21)，然后交换求和与积分符号的位置，得到：

$$\phi(t) = \frac{1}{q_{max}}\left[\frac{c_0}{2}\int_{-\infty}^{t} i(\tau)\mathrm{d}\tau + \sum_{n=1}^{\infty} c_n \int_{-\infty}^{t} i(\tau)\cos(n\omega_0 t)\mathrm{d}\tau\right] \tag{23}$$

这些数学运算的对应的步骤在图 17.9 的左半边图中用图表示出。注意，这个方框图再一次包含着类似于一个超外差接收机所有的元素。那个归一化了的噪声电流是一个宽带“射频”信号，其傅里叶分量经历了同时的下变频(即乘法运算)，这是用 ISF 的振荡频率的谐波成分充作“本振”信号来进行的。这里，重要的是要记住如果一个乘数是常数的话，乘法运算也是一个线性操作，我们现在所处的情况正是如此。这些乘法运算的相对贡献是由 ISF 的傅里

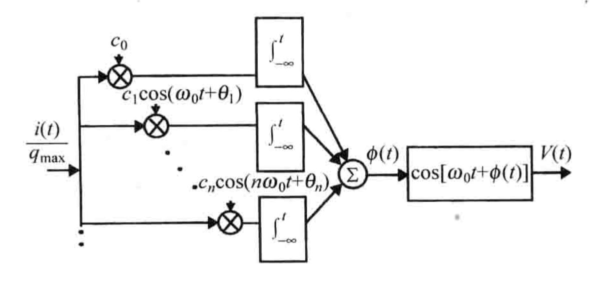


图 17.9　ISF(冲击激励函数)分解的等效系统图

叶系数决定的。一旦ISF的傅里叶系数被确定之后(通常通过仿真得到)，表达式(23)就允许我们来计算由一个任意的注入到系统中的噪声电流引起的过剩相位。

我们已经注意到一个常见的现象是注入一个处于某一个频率下的非线性系统的信号(即是噪声)可能会产生新的频率成分。现在说明一个线性但时变的系统也能呈现出定性类似的行为，就像为早先提到的超外差的图像所示的那样。为了清楚地说明这个性质，考虑将一个频率接近于系统振荡频率 m 整数倍的正弦电流注入进系统，亦即

$$i(t) = I_m \cos[(m\omega_0 + \Delta\omega)t] \tag{24}$$

其中，$\Delta\omega \ll \omega_0$。将式(24)代入式(23)，并且注意到除了 $n = m$ 的项，其他项对积分的贡献都可以忽略不计，于是推导出如下的近似式：

$$\phi(t) \approx \frac{I_m c_m \sin(\Delta\omega t)}{2q_{max}\Delta\omega} \tag{25}$$

这样，$\phi(t)$的频谱由在$\pm\Delta\omega$处两个相同的边带组成，尽管注入信号发生在 ω_0 的某个整数倍附近。这个结论是理解振荡器中噪声演变的基础。

遗憾的是，我们仍未完成需要做的事：式(25)使得我们可以推测出 $\phi(t)$的频谱，但最终想得到的是振荡器输出电压的频谱，这与 $\phi(t)$不完全是一回事。然而，这两个量的确通过实际的输出波形联系在了一起。为了说明我们所说的这种联系是什么，来考虑一个特定的例子，那里输出可以近似为正弦信号，因而有 $v_{out}(t)=\cos[\omega_0 t+\phi(t)]$。可以认为这个方程是一个由相位到电压的转换器，它的输入是相位，输出是电压。这个转换过程本质是非线性的，因为它包含了对一个正弦波的相位调制。

让我们来进行这个相位到电压的转换，并且假设小的幅度的扰动，我们发现单频信号的注入导致表达式(25)在载波两边对称地产生两个功率相等的边带

$$P_{SBC}(\Delta\omega) \approx 10\log\left(\frac{I_m c_m}{4q_{max}\Delta\omega}\right)^2 \tag{26}$$

注意这个幅度的依赖关系是线性的(式中的取平方运算仅仅是反映了我们这里是在与功率这个量打交道这个事实)。这个关系式已经为一个十分广泛数量的实际振荡器在实验上得到证实。

这个结论可以被推广到白噪声源的一般情形：

$$P_{SBC}(\Delta\omega) \approx 10\log\left[\frac{(\overline{i_n^2}/\Delta f)\sum_{m=0}^{\infty} c_m^2}{4q_{max}^2\ (\Delta\omega)^2}\right] \tag{27}$$

式(26)和式(27)一起指出了噪声信号通过上变频和下变频都被转换到载波附近的噪声，如图17.10所示。这张图总结了前面那些方程告诉我们的：所有整数倍载波频率附近的噪声都被变换到载波频率自身的附近。

直流附近的噪声被上变频，并且被乘以权重系数 c_0，故 $1/f$ 噪声变成载波附近的 $1/f^3$ 噪声；载波附近的噪声仍在载波附近，被乘以权重 c_1；载波的高阶整数倍附近的白噪声被下变频到载波附近，转换为 $1/f^2$ 区域的噪声。注意$1/f^2$ 的形状是从冲击型的噪声输入引起的相位阶跃变化所隐含的积分造成的。由于积分(即使是时变的)使得白色电压或电流谱具有一个 $1/f$ 的特征，功率谱密度就成了$1/f^2$ 形状。

从图17.10可以清楚地看出，如果能使得各种系数 c_n最小化(通过使ISF最小化来达

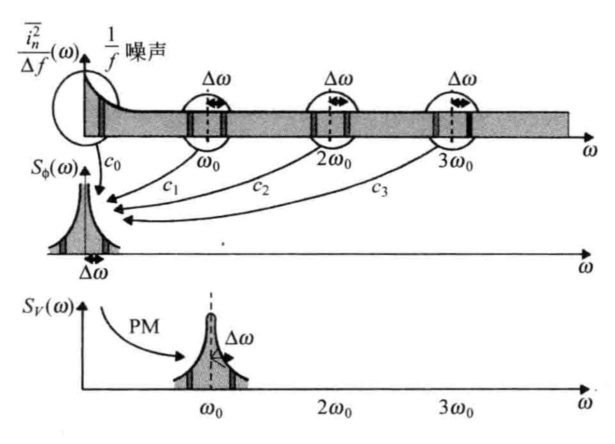


图 17.10 电路噪声到相位噪声的演变

到)，就能使相位噪声最小化。为了定量地强调这一点，可以用 Parseval 定理写出：

$$\sum_{m=0}^{\infty} c_m^2 = \frac{1}{\pi}\int_0^{2\pi} |\Gamma(x)|^2 \mathrm{d}x = 2\Gamma_{\mathrm{rms}}^2 \tag{28}$$

这样在 $1/f^2$ 区域的频谱可以表达为

$$L(\Delta\omega) = 10\log\left[\frac{(\overline{i_n^2}/\Delta f)\Gamma_{\mathrm{rms}}^2}{2q_{\max}^2\ (\Delta\omega)^2}\right] \tag{29}$$

这里 Γ_{rms}是 ISF 的均方根值。如果其他参数都保持相等，减小 Γ_{rms}可以减小在所有频域下的相位噪声。表达式(29)是$1/f^2$ 区域内的严格表达式，也是这个噪声模型的一个关键结果。注意，在这个表达式中没有经验曲线拟合参数。

表达式(29)的其他特性包括允许我们定量地研究 $1/f$ 噪声是如何通过上变频成为载波附近的相位噪声。载波附近的噪声对于通道间距很窄的通信系统显得尤为重要。事实上，实际的通道间距常常为相位噪声所限制。遗憾的是，采用线性时不变(LTI)模型无法正确地预测载波附近的相位噪声。

如果用这个新的模型，这个问题就消失了。特别是，假设电流噪声的行为在 $1/f$ 区域内的行为如下：

$$\overline{i_{n,1/f}^2} = \overline{i_n^2} \cdot \frac{\omega_{1/f}}{\Delta\omega} \tag{30}$$

这里 $\omega_{1/f}$是 $1/f$ 的拐角频率。用式(27)，得到下述在 $1/f^3$ 区域的噪声表达式

$$L(\Delta\omega) = 10\log\left[\frac{(\overline{i_n^2}/\Delta f)c_0^2}{8q_{\max}^2\ (\Delta\omega)^2} \cdot \frac{\omega_{1/f}}{\Delta\omega}\right] \tag{31}$$

因此 $1/f^3$ 的拐角频率为

$$\Delta\omega_{1/f^3} = \omega_{1/f}\frac{c_0^2}{4\Gamma_{\mathrm{rms}}^2} \approx \omega_{1/f}\left(\frac{\Gamma_{\mathrm{dc}}}{\Gamma_{\mathrm{rms}}}\right)^2 \tag{32}$$

由此我们可以看到，$1/f^3$ 的相位噪声拐角频率并不必然与 $1/f$ 器件或电路的噪声拐点一致，而且一般说来这个值要低一些。事实上，由于 Γ_{dc}是 ISF 的直流分量，因此存在着一种可能将 $1/f^3$ 拐角频率大幅度地降低。ISF 是波形的函数，因而潜在地说是可以由设计者来控制的，这通常是通过调整上升与下降时间的对称性。这个结果由线性时不变(LTI)模型是

得不到的，这也是线性时变(LTV)模型提供的最有用处的内在理解之一。这个结果对于 $1/f$ 噪声性能很差的器件工艺具有特别重要的意义，如 GaAs MESFET 与 CMOS 工艺。有关如何可以利用这个观察得到的结论来进行电路设计的例子将在 17.5 节给出。

再有一个特别有用的理解涉及到周期性稳态的噪声源。就如前面所讨论的，在许多振荡器中的噪声源并不能看作是稳态的。一个典型的例子是在 FET 中的标称漏极白噪声电流或双极型晶体管中的散弹噪声。噪声电流是偏置电流的函数，而后者随着振荡波形周期性且十分显著地发生变化。这个 LTV 模型可以很容易地包括周期性白噪声源，因为可以将这样的噪声源看作是一个稳态的白噪声源与一个周期函数的乘积[6]：

$$i_{\mathrm{n}}(t) = i_{\mathrm{n0}}(t) \cdot \alpha(\omega_0 t) \tag{33}$$

其中，i_{n0} 是一个稳态的白噪声源，其峰值等于周期性稳态噪声源的峰值，而噪声调制函数(noise modulation function，NMF)$\alpha(x)$是一个无量纲的周期函数，其峰值是 1，如图 17.11 所示。将此式代入式(21)，可以将一个周期性稳态噪声处理成一个稳态噪声源，这样做的条件是定义一个如下的等效 ISF

$$\Gamma_{\mathrm{eff}}(x) = \Gamma(x) \cdot \alpha(x) \tag{34}$$

图 17.12 对一个 Colpitts 振荡器给出了在一个周期内的 $\Gamma(x)$、$\alpha(x)$以及 $\Gamma_{\mathrm{eff}}(x)$。$\Gamma(x)$的准正弦波形状或许是可以预见到的，这是基于之前研究过的理想 LC 振荡器的 ISF，在这种振荡器中输出电压与 ISF 在形状上是大致一样的，但相位是正交的。在大部分时间中，NMF 几乎为零，这是与在 Colpitts 电路中晶体管工作在 C 类状态是一致的，晶体管在一个相对较窄的时间窗口内补充损耗的谐振回路能量，如同函数 $\alpha(x)$的形状所隐指的那样。这两个函数的乘积 $\Gamma_{\mathrm{eff}}(x)$有比 $\Gamma(x)$小得多的均方根值，这清楚地说明了这个振荡器所利用的周期稳态性。

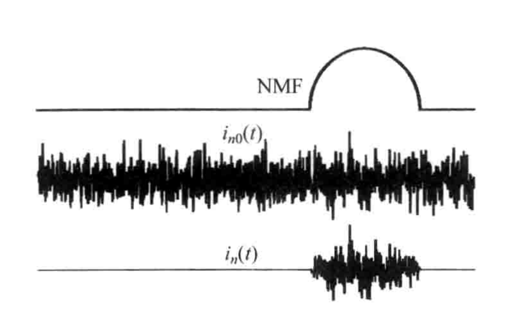


图 17.11　作为稳态噪声和 NMF 乘积的周期性稳态噪声

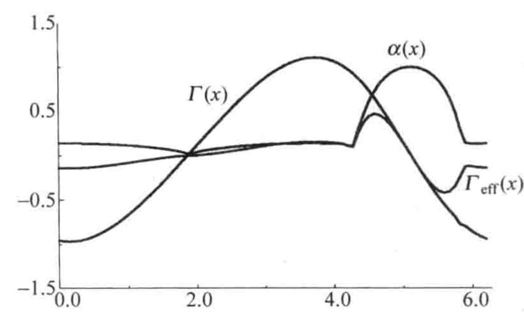


图 17.12　周期性稳态的容纳

这个例子强调了周期稳态性，因此是容易纳入到我们已经建立的框架内。只要在所有的方程中都用 Γ_{eff}，前面的结论都不会发生改变[7]。

---

⑥ W. A. Gardner，*Introduction to Random Processes*(随机过程导论)，McGraw-Hill，New York，1990.

⑦ 这个公式不一定能适用如果外部的周期性稳态这是源被引入，例如可能在注入锁定振荡器(injection-locked oscillator)里就是这样的情形。要知道详细的讨论，见 P. Vanassche，*et al.*，"On the difference between two widely publicized methods for analyzing oscillator phase behavior"(论两个广泛被讨论的分析振荡器相位行为的方法之间的差别)，*Proc. IEEE/ACM/ICCAD*，session 4A，2002.

在已经确定了影响振荡器噪声的因素之后，可以明确地说出设计一个好的振荡器需要满足的条件。首先，与 LTI 模型所指明的相一致的，信号功率与谐振器的 Q 值需要最大化，这是假设在其他的因素都保持相同的条件下。此外，注意到谐振回路是一定需要一个有源器件来补偿其中的损耗的，而有源器件总是会产生噪声。同时要注意，ISF 告诉我们在一个振荡周期中存在着敏感与不敏感的时刻。在几乎是无穷多种有源器件能为谐振回路补充能量的方式中，一个最好的策略就是在某一时刻一次性地补充所有需要的能量，而在这一时刻 ISF 具有最小值。这样，在一个理想的振荡器中，晶体管在几乎所有的时间上都处于关断状态，只是周期性地在每个周期信号处于峰值时注入一个冲击电流。实际的振荡器能在多大程度上趋近于这种行为，将在很大程度上决定它们的相位噪声性质。因为 LTI 理论认为所有的时刻都是同样重要的，故而这种理论是不能预测到这个重要的结果。

这个采用冲击能量恢复的法子实际上已被用了几个世纪了，但是是在不同的场合。在机械钟表里，一个被称为擒纵机构的装置管理着能量从弹簧到摆锤的转移。这个擒纵机构强迫这个能量转移是以冲击的形式发生，而且只在精确的算好的时刻发生(与摆锤达到最大速度的时刻一致)，选择这个时刻是为了使冲击对振荡周期的扰动最小化。尽管这个具有重要历史性比拟的装置已经是存在了几百年，却是被凭借直觉设计与反复试验得到的，直到 1826 年才被天文学家 Royal George Airy⑧ 在数学上分析说明了。当然，这个机构与更为广泛的电子振荡领域的联系只是在近年来才被认识到。

最后，性能最好的振荡器要具有对称性质，这种对称性可以得到小的 Γ_{dc} 以使 $1/f$ 噪声的上变频最小化。在研究了一些近场相位噪声的附加特性后，我们在下一节考虑几个电路例子，以说明在实际中如何达到这些目的。

近场(close-in)相位噪声

根据到目前为止的讨论，我们会期待 $S_\phi(\omega)$ 的频谱有一个反比于频率立方的近场行为。亦即，当趋近载波频率时谱密度没有限制地增大。然而，大多数测量并没有显示出这个行为，同时这个没有得到的实验证实常常被错误地解释成某个新现象的结果，或是在 LTV 理论中存在缺陷的证据。因此，值得我们花一点时间来详细地考虑这个问题。

LTV 这个理论只认定 $S_\phi(\omega)$ 是不受限制地增加。而大多数的“相位”噪声测量实际上测的是振荡器的输出电压的谱。也就是说，经常所测的实际上是 $S_V(\omega)$。在这样的情况下，当偏离频率趋近于零时，输出的谱不会表现出不受限制的增长，从而反映了这样一个简单的事实，即一个正弦函数甚至对于不受限制的自变量也是一个有限的量。这个限制引起当频率趋近于载波时测量得到的频谱变平坦，而得到的形状是洛伦兹峰(Lorentzian)形⑨，如图 17.13 所示。

---

⑧ G. B. Airy, “On the disturbances of pendulums and balances, and on the escapements”(论摆锤的扰动与平衡，及擒纵机构)，*Trans. Cambridge Philos. Soc.*, v. 3, pt. I, 1830, pp. 105-128。本书的作者十分感谢 Byron Blanchard 指出了这个文献。

⑨ W. A. Edson, “Noise in oscillators”(振荡器中的噪声)，*Proc. IRE*, August 1960, pp. 1454-1466。也参见 J. A. Mullen, “Background noise in nonlinear oscillators”(非线性振荡器中的背景噪声)，*Proc. IRE*, August 1960, pp. 1467-1473。洛伦兹形状与一个单一极点的低通滤波器的功率响应是一样的。只不过当你说“洛伦兹形”，听起来更令人印象深刻些。

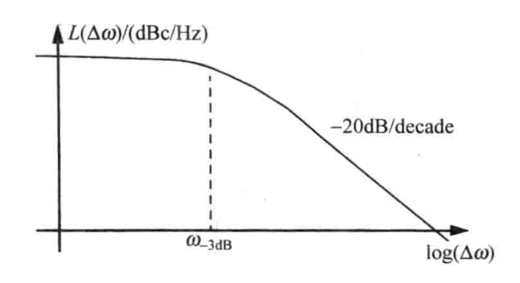


图 17.13 洛伦兹峰(Lorentzian)形频谱

这个－3dB 拐角可能观察到，也可能观察不到，这取决于进行测量的具体细节。如果用频谱分析仪，这个拐角在通常情况下是可以观察到的。如果有一个理想的鉴相器与一个锁相环来将 $\phi(t)$ 的频谱下变频，并且直接测量，那么完全不会观察到平坦区。采用实际的鉴相器(其必然具有有限的相位检测范围)，通常会观察到一个－3dB 拐角，但是这个拐角的精确值现在就是测量设置的函数了，测量的结果将不再反映振荡器的本征谱性质。这个缺乏在测量技术上的一致性在过去一直是很大的造成混乱的来源。

17.5 几个电路实例——*LC* 振荡器

在已经推导了在低的以及中等的频偏下相位噪声的表达式之后，将已经获得的内在认识应用到实际的振荡器中，这样做是有指导意义的。首先研究常用的 Colpitts 振荡器及其有关的波形(见图 17.14 和图 17.15)。一个重要的特征是漏极电流只在一个短的间隔中存在，这个间隔与最不会惹事的时刻(即那些谐振回路电压的峰值处)相一致。这个所对应的极佳的相位噪声性质解释了这个结构为什么被广泛采用。人们早就知道最好的相位噪声特性出现在某个窄的抽头比范围内(比如电容比 C_2/C_1 为 3∶1 或 4∶1)，但是在 LTV 理论出现之前，没有理论基础可以解释特定的最佳值。

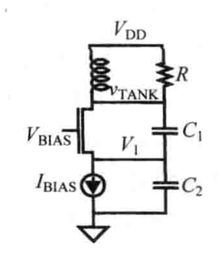


图 17.14 Colpitts 振荡器(简化了的)

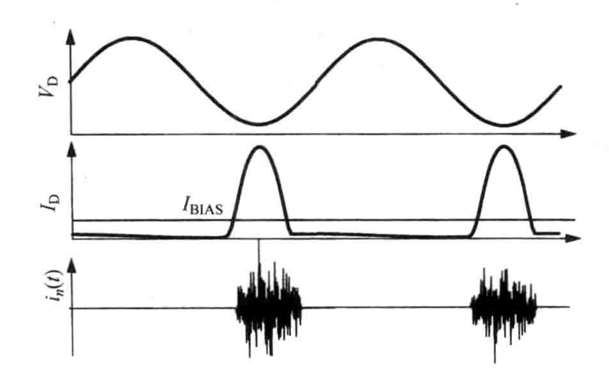


图 17.15 Colpitts 振荡器的近似的增量谐振回路电压与漏极电流

从图 17.15 可以看出漏极电流噪声的周期稳态特征是十分明显的。因为当 ISF 相对来说是小的时候，噪声是最大，有效 ISF(即 ISF 与噪声调制函数的乘积)要比 ISF 小得多。

LTI 与 LTV 这两个模型都指出了将信号幅度最大化的价值。为了避免受电源电压或击穿电压的限制，可以采用带抽头的谐振器以将谐振器两端的电压摆幅与器件电压的限制去耦。采用这种方法的一个结构是 Clapp 对 Colpitts 振荡器(重新画在图 17.16 中)的改动。采用带抽头谐振器的振荡器的差分实现形式也已出现在文献中[10]。这些类型的振荡器或者是 Clapp 结构或者是它的对偶(带有抽头的电感而不是电容分压器)。Clapp 结构在电源电压变得越来越低时，变得越来越有吸引力，在电源电压低的场合，传统的谐振器的连接致使摆幅受 V_{DD} 限制。采用抽头的方法可以使信号的能量即使在电源电压低时也可以保持在高水平。

采用 LTV 模型对相位噪声的预测通常是对双极型振荡器更为准确些，这是由于有较好的器件噪声模型可以利用。在 Margarit 等著的文章中(见脚注 10)，冲击响应的建模(见 17.8 节)被用来决定对在振荡器中的各种噪声源的 ISF，这个知识然后被用来优化一个差分双极型 VCO 的噪声表现。这个振荡器的一个简化了的示意图在图 17.17 中给出。

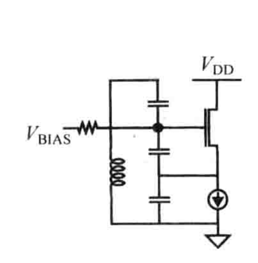


图 17.16　Clapp 振荡器

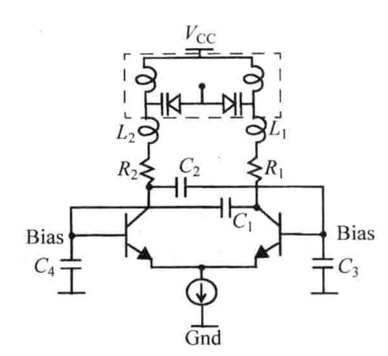


图 17.17　Margarit 等文(1999)中的压控振荡器的简化示意图

带有抽头的谐振器被用来增加谐振回路的信号功率 P_{sig}。电容抽头比的最优值大约是 4.5(相应的电容比为 3.5)，这是基于将噪声源的周期性、稳态性考虑在内的模拟。特别是，这个模拟考虑了因基区的扩展电阻(spreading resistance)的噪声贡献，及每个晶体管的集电极散弹噪声，还有因谐振回路元件的电阻性损耗而引起的噪声。分别对应于核心振荡器中的晶体管贡献的散弹噪声以及偏置源噪声的 ISF 在图 17.18 及图 17.19 给出(ISF 是取自 Margarit 文，它们是通过如在 17.8 节中描述的时域里直接计算出来的)。在图中可以看到，尾电流的噪声有一个 ISF，其周期是振荡器周期的一半，这是由这个电路的差分拓扑结

[10] J. Craninckx and M. Steyaert, "A 1.8GHz CMOS low-phase-noise voltage-controlled oscillator with prescaler"(一个带有预分频器的 CMOS 1.8GHz 低噪声压控振荡器), *IEEE J. Solid-State Circuits*, v. 30, no. 12, December 1995, pp. 1474-1482。也参见 M. A. Margarit, J. I. Tham, R. G. Meyer, and M. J. Deen, "A low-noise, low-powerVCO with automatic amplitude control for wireless applications"(一个带有自动幅度控制的用于无线应用的低噪声、低功耗压控振荡器), *IEEE J. Solid-State Circuits*, v. 34, no. 6, June 1999, pp. 761-771。

构(尾电压的波形含有一个频率为振荡器频率两倍的成分)而引起的。下面这个观察是有价值,即尾噪声对相位噪声的贡献因此只在振荡频率的偶数倍时发生。如果尾电流在被供给振荡器的核心电路之前经过一个低通(或单边带通)滤波器,那么这个尾电流所贡献的噪声会被大大降低,噪声的降低达到 10dB 或更多已经见于报道[11]。只有尾电流的 $1/f$ 噪声会对噪声作出贡献。各个 *ISF* 被用来计算每一个相应的噪声源的贡献,然后这些贡献被加起来。

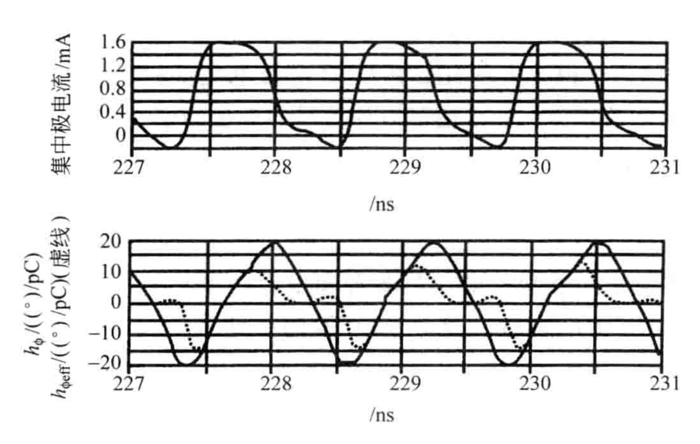


图 17.18　电路中每一个核心晶体管的散弹噪声的 ISF(摘自 Margarit 等,1999)

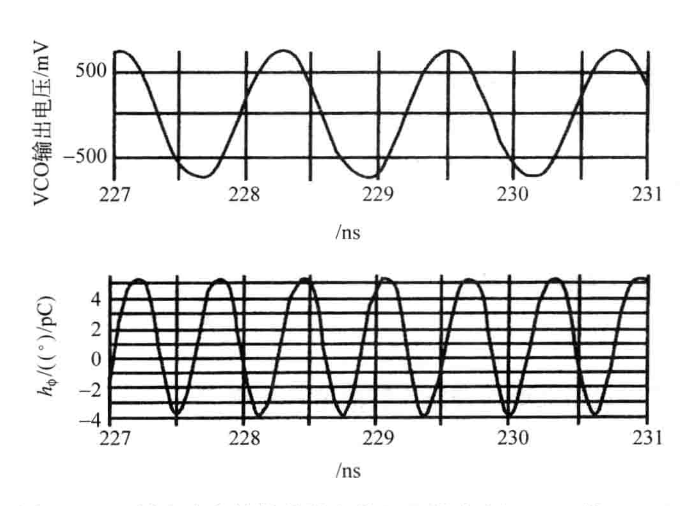


图 17.19　尾电流中的散弹噪声的 ISF(摘自 Margarit 等,1999)

采用这个拓扑电路对 $1/f$ 噪声下变频的降低可以清楚地在图 17.20 中看出,那里给出了预测及测量到的 $1/f^3$ 的拐角频率为 3kHz,与一个独立的 $1/f$ 噪声拐角为 200kHz 的比较。注意,采用当时是最新的商用模拟工具 Spectre(EDA 公司 Cadence 的产品)对这个例

[11] A. Hajimiri and T. Lee 的书 *The Design of Low-Noise Oscillators*(低噪声振荡器的设计,Kluwer, Dordrecht, 1999)描述了一个简单的并联电容连接尾电路节点与地。E. Hegazi, *et al.*, "A filtering technique to lower oscillator phase noise"(一个降低振荡器噪声的滤波技术,*ISSCC Digest of Technical Papers*, February 2001)采用了一个在尾电流源与共源节点之间的并联谐振槽路来取得 10dB 相位噪声的降低。

子没有能在图中所示的偏离频率范围内确认出$1/f^3$拐角，结果是当 100Hz 偏离频率时，低估了大约有 15dB。在$1/f^2$区域路测量到的相位噪声也是与 LTV 模型预测的十分相符。比如说，在 100kHz 频偏的预测值－106.2dBc/Hz 与测量到的值－106dBc/Hz 几乎没有差别。作为最后一个说明，值得提到这个特别的 VCO 设计用了一个单独的自动幅度控制环路，这个环路允许独立的稳态与启动状态的优化，这样的功能对取得良好的相位噪声表现是有帮助的。

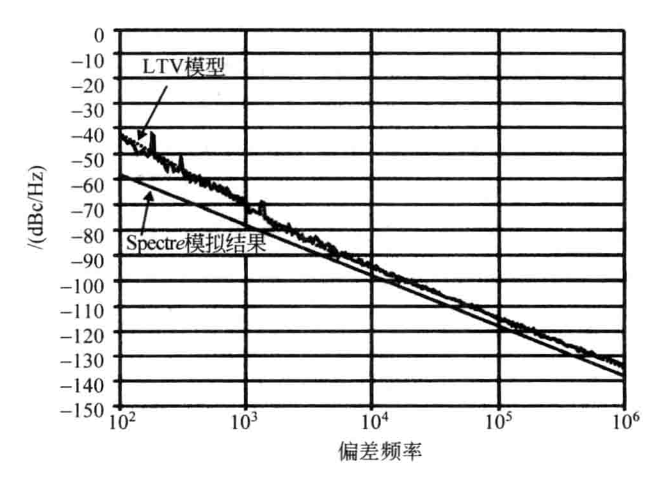


图 17.20　在 Margarit 等文(1999)中测量到和预测的压控振荡器的相位噪声

就如已提到过的，LTV 理论提供的一个关键性的内在见解涉及对称性的重要性，在前面这个例子中，这个对称性造成的一些效应是部分明显可见的。更充分地利用这个知识的结构是一个对称负阻振荡器，如图 17.21 所示[12]。这个结构从哪一方面来说都不是新的，但是对于它的对称性性质的理解却是新的。这里，正是那半个电路的对称性是重要的，因为在这两个半电路中的噪声在最不利的情况下也只是部分相关联的。通过恰当地选择 PMOS 与 NMOS 器件的相对宽带来使对每半个电路的 ISF 直流值(Γ_{dc})最小化，我们可以使得$1/f$噪声的上变频为最小。通过以这样的方式来利用对称性，$1/f^3$拐角频率可以变得极其小，即使是在器件的$1/f$噪声拐角是大的情形下(这对 CMOS 而言典型的情况就是如此：$1/f$噪声拐角大)。进一步说，这四个晶体管桥式的接法使得信号的摆幅可以更大，又有利于相位噪声的改善。作为所有这些因素的结果，对于 1.8GHz 的载波，在 600kHz 的频偏下，可以取得相位噪声值为－121dBc/Hz，这个结果是在用低 Q 的在片螺旋电感，功耗 6mW，0.25 μm CMOS 工艺下得到的(见脚注 12)。这个结果可以与采用双极型技术能达到的指标相媲美，这可以从与 Margarit 等文(1999)中的双极型例子的比较见到。只需在功耗上稍微增加，同样的振荡器的相位噪声可以满足 GSM1800 指标的要求。

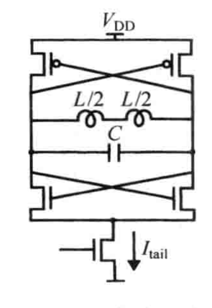


图 17.21　简单对称负阻振荡器

[12] A. Hajimiri and T. Lee, "Design issues in CMOS differential LC oscillators"(CMOS 差分 LC 振荡器的几个设计议题), *IEEE J. Solid-State Circuis*, May 1999.

17.6 振幅响应

虽然近场边带主要由相位噪声决定，远场边带却是大大地受幅度噪声的影响。不像感生的过剩相位，由于一个电流冲击造成的过剩幅度 $A(t)$ 随着时间而衰减。这个衰减总是存在于实际的振荡器中的幅度恢复机制的直接结果。这个过剩幅度可能衰减起来非常慢（比如，在一个具有高 Q 值谐振回路的谐波振荡器）或非常快（比如在一个环振）。某些电路甚至可能呈现出一个欠临界阻尼的二阶幅度响应。幅度控制机理的详细的动态表现会在噪声谱的形状是有直接的效应。

在如图 17.5 所示的理想 LC 振荡器的环境下，一个面积为 Δq 的电路冲击会在电容上的电压感生出一个瞬时的变化，这又会造成振荡器幅度的变化，而幅度的变化是依赖于注入的瞬间（如图 17.6 所示）。这个幅度的变化正比于瞬间归一化了的电压变化 $\Delta V/V_{\max}$，条件是注入电荷要小，以满足 $\Delta q \ll q_{\max}$

$$\Delta A = \Lambda(\omega_0 t)\frac{\Delta V}{V_{\max}} = \Lambda(\omega_0 t)\frac{\Delta q}{q_{\max}}, \quad \Delta q \ll q_{\text{swing}} \tag{35}$$

其中幅度冲击灵敏度函数 $\Lambda(\omega_0 t)$ 是一个周期函数，它决定了波形中的每一点对一个冲击的灵敏度。类似于对相位响应的处理方法，幅度冲击响应可以被写成

$$h_A(t,\tau) = \frac{\Lambda(\omega_0 t)}{q_{\max}} d(t-\tau) \tag{36}$$

式中 $d(t-\tau)$ 是定义过剩幅度如何衰减的函数。图 17.22 给出了两个假想的例子：带有过阻尼响应的低 Q 振荡器与带有欠阻尼幅度响应的高 Q 振荡器的 $d(t)$。

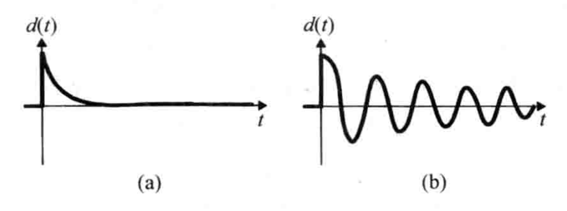


图 17.22 过阻尼与欠阻尼的幅度响应

就如对相位响应的求值，这里引用了一个小信号线性近似。又一次，我们并没有忽略幅度控制的基本的非线性本质，只是利用了这样一个优势，即幅度噪声必然是足够小以使得对任何一个值得我们来分析的振荡器可以采用小信号近似。在不失一般性的前提下，假设大多数振荡器的幅度限制系统可以被近似为一阶或二阶行为，这是又一次对小的扰动而言。因此，函数 $d(t-\tau)$ 在典型的情况下或者是一个衰减的指数函数，或者是一个被阻尼的正弦波。

对一个一阶系统，

$$d(t-\tau) = \mathrm{e}^{-\omega_0(t-\tau)/Q} \cdot u(t-\tau) \tag{37}$$

因此，对一个任意的输入电流 $i(t)$，过剩幅度响应是由一个叠加积分给出的：

$$A(t) = \int_{-\infty}^{t} \frac{i(\tau)}{q_{\max}} \Lambda(\omega_0 t) \mathrm{e}^{-\omega_0(t-\tau)/Q} \mathrm{d}\tau \tag{38}$$

假若 $i(t)$ 的功率谱密度是一个白噪声源，那么幅度噪声 $A(t)$ 的输出功率谱可以被证

明是

$$L_{\text{amplitude}}\{\Delta\omega\}=\frac{\Lambda_{\text{rms}}^2}{q_{\text{max}}^2}\cdot\frac{\overline{i_n^2}/\Delta f}{2[\omega_0^2/Q^2+(\Delta\omega)^2]} \tag{39}$$

其中，Λ_{rms}是$\Lambda(\omega_0 t)$的均方根值。如果L_{total}是测量得到的，那么$L_{\text{amplitude}}$与L_{phase}的和会被观察到，因而相位噪声谱在ω_0/Q处会有平台，如图 17.23 所示。同时也注意到幅度响应的重要性在很大程度上取决于Λ_{rms}，而这个量又取决于电路拓扑。

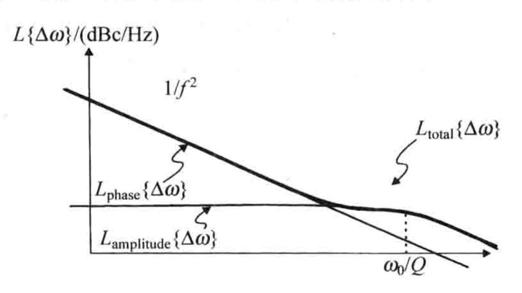


图 17.23　过阻尼幅度响应的相位、振幅及总的边带功率

作为对幅度控制动态行为的最后一个说明是，一个次阻尼的响应会产生一个谱，它在ω_0/Q附近会呈现某个峰值。

17.7　总结

从线性时不变(LTI)模型得到的内在理解是简单的而又直觉上令人满意的：那就是我们应该让信号幅度与谐振器的 Q 值最大化。一个额外的、隐式的内在见解是环路的相移一般来说必须安排至振荡发生在谐振回路的中心频率附近。这样做了，偏离中心频率的谱分量就会被谐振器最大限度地衰减掉。

为 LTV 模型提供的更深层次的内在理解是谐振器的能量应该在 ISF 的最小值处以冲击形式，而不是在整个周期内均匀地得以补充，而且有效 ISF 的直流值应该尽可能被做成零，这样就可以抑制 1/f 噪声被上变频至近场相位噪声。这个理论也指出，环振宽带噪声性能不佳的缺点可以利用它们潜在的抑制共模衬底与电源噪声的优良能力来弥补。

17.8　附录　有关仿真的说明

除了最简单的振荡器外，通常是不可能得到 ISF 的精确解析表示式的。在脚注 4 中引用的参考文献中提及了各种近似方法，但是只有直接计算时变冲击响应的方法算得上是通用准确方法。在这个直接方法里，冲击激励扰动了振荡器，然后稳态相位扰动被测量。冲击相对于未被扰动的振荡器过零时刻的瞬间然后被逐个变动，这样的模拟被重复直至冲击“走遍”了整个周期。

这个冲击必须要有足够小的幅度以确保线性假设的成立。正像对一个放大器的阶梯响应在阶梯输入的幅度具有任意大的值不能被恰当地求值，我们必须适当地选择冲击的面积而不是盲目地采用某个固定的值(例如 1C)。假若不确信所选的冲击是否有恰当的大小，必

须专门进行线性性的测试。这个测试是通过改变冲击的大小，然后检查响应的幅度是否改变同样的比例。

最后，一直存在着这样一个关于 LTV(线性时变)理论是否恰当地包含了某些振荡器表现出来的幅度至相位的转换现象的模糊理解。只要线性性是成立的，这个 LTV 理论的确覆盖了 AM 到 PM 的转换，这是假设一个精确的 ISF 已被得到。之所以这么说是基于这样一个事实，即源于幅度变化的一个振荡器的相位变化会出现在振荡器的冲击响应中。然而，有一个细微之处会从这两个机理产生的边带间的关系中出现。将来自这两个源的贡献加起来会导致幅度不相等的边带，这不同于单个 AM 和 PM 各自所特有的纯对称边带。

第18章 相位噪声测量

18.1 引言

振荡器相位噪声的重要性来源和在有关振荡器和相位噪声的相关章节中已经阐述过。现在我们来研究如何测量它。甚至比测量放大器的噪声系数更难，相位噪声的测量很容易被来自众多方面的误差来源而失败，因此明白什么是这些误差的来源(及如何减小它们)就变得十分必要。使这个任务变得更复杂的是许多参考文献给出了一些近似但是却没有明确地说明这些近似是什么。在本章中，我们将致力于指明这些近似，特别是着重在这些近似成立的区域。

遵循这本书的写作理念，除了给出在专业实验室里通常使用的相位噪声测量技术外，那些适用于周末工作的业余爱好者的测量方法也予以介绍。

18.2 定义与基本测量方法

正如第15章讨论的，所有真实的振荡器输出都呈现出某种相位和幅度的变化：

$$V_{out}(t) = V_m[1+\varepsilon(t)]\cos[\omega t+\phi(t)] \tag{1}$$

由于为 $\varepsilon(t)$ 和 $\phi(t)$ 所代表的不希望发生的幅度和相位调制，使得输出的频谱在不仅仅是振荡器的标称频率其他频率上有宽的分布能量。一种典型的 V_{out} 频谱因此看上去如图18.1所示，此图很清晰地显示了噪声功率随频率的变化。通常来讲，所显示的频谱是幅度和相位噪声的结合，但是我们经常主要关注导致相位变化的那部分噪声。其原因是在所有实际的振荡器里总是存在某种形式的幅度限制机制。但是，并没有类似的限制相位变化的机制存在(因为无法让振荡器确立真正的时间起点)，因此相位噪声有占主导作用的倾向。当然，我们也并不能以此推论说幅度噪声是不相关的：幅度噪声能引起各种各样的奇怪的行为，尤其是在频率远离载波频率的地方[①]。话是这么说，相位噪声是绝大多数振荡器设计者主要关心的问题，因此也是我们的关注重点。

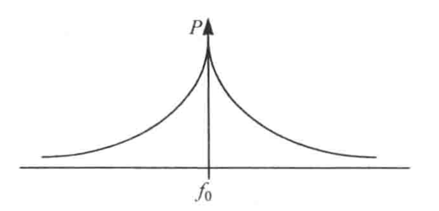


图18.1 一个理想振荡器的理想化了的双边带输出频谱

下面相位噪声这个定义通常被用来表征振荡器：

$$\mathcal{L}(\Delta f) \stackrel{\text{def}}{=} \frac{P_{SSB}(1\text{Hz})}{P_{sig}} \tag{2}$$

---

① 对那些想看众多的积分式子的读者，*The Design of Low-Noise Oscillators*(低噪声放大器设计)(A. Hajimiri and T. Lee, Kluwer, Dordrecht, 1999)这本书详细讨论了器件的噪声是如何变成振荡器的噪声的。

这里$P_{SSB}(1\text{Hz})$是偏移载波频率为Δf处的输出相位噪声功率密度，P_{sig}代表载波自身的功率。因此，$\mathcal{L}(\Delta f)$，被称为手写L，代表被载波功率归一化之后的功率密度。习惯上用分贝形式表示为 dBc/Hz（相对于载波而言），尽管其中的每赫兹部分是对对数里面的变量而言，不是对整个对数本身。采用这种写法虽然缺少严密性，但是是个习惯用法。只是不要被误导：带宽翻倍不意味着用分贝表示的噪声翻倍。

如果幅度噪声被认为是可以忽略的（这是一个很大的假设），那么相位噪声能够从频谱分析仪直接读出来。要这么做，首先将在希望的偏移量Δf处显示的噪声功率除以噪声带宽的分辨率设置，将其归一化到 1Hz 带宽。在很多的频谱分析仪中，其内部的设置带宽精度的中频滤波器被同步调谐（即，一串相同的滤波级被调到同一中心频率），因此就产生了一个高斯形状的响应。它们的噪声带宽因此近似用 1.2 倍于 −3dB 的带宽精度。最终用信号功率归一化之后就得到了$\mathcal{L}(\Delta f)$，如图 18.2 所示。

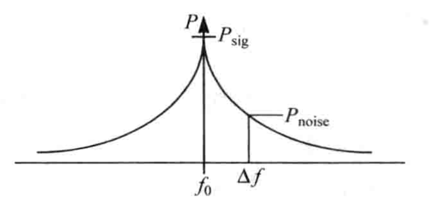


图 18.2　从频谱分析仪的显示来估计 $L(\Delta f)$

作为一个专门的数值例子，假设载波功率P_{sig}为 50mW (17dBm)，而P_{noise}为 2nW(−57dBm)，测量的分辨率带宽是 100Hz(20dBHz)，频率偏移量为 600kHz。那么相位噪声近似等于

$$\mathcal{L}(\Delta f)=\frac{P_{SSB}(1\text{Hz})}{P_{sig}}=\frac{2\text{nW}}{50\text{mW}}\cdot\frac{1}{1.2(100\text{Hz})}=3.33\times10^{-10}\ \text{Hz}^{-1} \tag{3}$$

或者大约为 −95dBc/Hz。再一次假设幅度噪声在这个频率偏移下可以忽略，并且用一个高斯滤波器来设定分辨率带宽。许多频谱分析器包含了一个用于进行自动归一化设定的模式：只需把游标移动到需要的偏移，仪器就会计算和显示它的近似的相位噪声$\mathcal{L}(\Delta f)$。（再一次指出，所近似的部分是我们有关任何幅度噪声是可忽略的假设。）

注意基于频谱分析器测量的灵敏度是被与载波的存在有关的动态范围问题所约束的。通过增加增益来提高灵敏度的努力很容易导致载波功率使分析仪过载（一个典型的频谱分析仪的赝空动态范围大约 70dB 左右）。在待测器件（DUT）和分析仪之间插入一个陷波滤波器来移除载波可以减轻这个动态范围问题，但这是以在载波附近引入频谱失真为代价的。如果这个失真可以接受（或者至少可以被表征，然后除去），那么陷波滤波器就是一种增加频谱分析技术使用范围的简单方法。

另一种表征相位噪声的方法是报告相位噪声本身的实际功率谱密度：

$$S_{\phi}(\Delta f)\stackrel{\text{def}}{=}\overline{\phi^2}/B \tag{4}$$

其中，B是测量进行的噪声带宽。在最大相位偏移是小（$|\phi|\ll1$）的情况下，可以采用这样一个近似，即相位“调制”功率的一半处于一个边带，而剩余的在另一个边带。在这个特殊情形下，可以做如下近似：

$$\mathcal{L}(\Delta f)\approx S_{\phi}(\Delta f)/2 \tag{5}$$

系数 2 简单地是因为$\mathcal{L}(\Delta f)$被定义为单边带量而产生的，而载波的相位调制结果是在双边带内的。

另一种偶尔使用的测量是频率噪声的功率谱密度。现在频率是相位的时间微分，因此乘上频率就是一个在频域里类似对时间的微分。因此，

$$s_{f}(\Delta f)\stackrel{\text{def}}{=}\overline{\Delta f^2}/B=f^2S_{\phi}(\Delta f)=2L(\Delta f) \tag{6}$$

其中(又一次)最后这个近似是基于足够小的相位偏移的假设。

18.3　测量技术

在提供了相位噪声的基本定义并且描述了一个简单直接的使用频谱分析器进行测量的方法后，现在考虑测量相位噪声的另外一些替代技术。

那个频谱分析仪的例子突显了动态范围的问题，这个问题是伴随任何一种寻求在载波信号存在情况下测量噪声都会发生的。正如先前所提到的，采用陷波滤波器能够减少这些问题，使得灵敏度增加。另一种替代方法是采用外差式结构来将载波附近的信号下变频至直流信号，这样就允许在没有干扰载波的情况下测量噪声。下面给出的三种方法描述了如何在测量中剔除载波的不同策略。

18.3.1　基于锁相环的监相器技术

既然想要表征相位的波动，那么最敏感的相位噪声测量方法是采用监相器作为测试装置的核心就不足为奇了。由于监相器的核心地位，了解其特征就是很重要了。

一个理想的监相器能够产生一个跟幅度完全无关而仅仅正比于两个输入相位(例如一个是从被测的器件，另一个是在仪器本身中产生的输入)的差的输出，见图 18.3。因为这个特性恰恰是想要的，墨菲定律保证不会得到那么理想的特性(玩笑话，译者注)。典型的情况是，采用一个模拟乘法器(图 18.4)来近似用作鉴相器。利用某些三角恒等式，可以把乘法器的输出表示为

$$AB\cos\omega t\cos(\omega t+\phi)=k_{\mathrm{d}}[\cos\phi-\cos(2\omega t+\phi)] \tag{7}$$

其中k_{d}是一个常数。

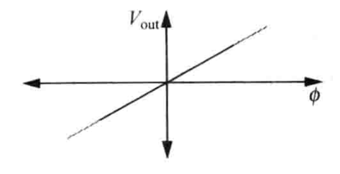


图 18.3　理想的鉴相器特性

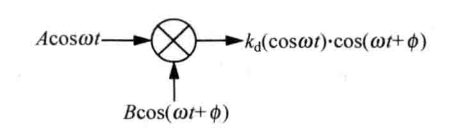


图 18.4　作为鉴相器的乘法器

注意乘法器的输出包括一个直流项和一个 2 倍频项。对于鉴相器，仅关注直流项因为它是唯一取决于相位差的函数。这个项恰恰是鉴相器的时间平均输出：

$$\langle AB\cos\omega t\cos(\omega t+\phi)\rangle=k_{\mathrm{d}}[\cos\phi] \tag{8}$$

如果我们画出平均输出作为相位角的函数，结果就会如图 18.5 所示。注意输出是周期性的于十分非线性的，即使一个周期内也是如此。由于输出是周期性的，很大的相位误差和很小的相位误差是区分不出来的。进一步注意到，鉴相器的灵敏度(即曲线的斜率)当相位之差为 0 时是零，而在输入相位之差为90°时，灵敏度达最大值。因此，为了使鉴相器

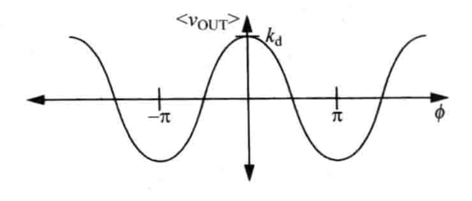


图 18.5　采用乘法器实现的鉴相器的输出与相位差的关系曲线

有用，应使得被测器件与参考的振荡器相位标称正交。

从图 18.5 可以看出，鉴相器的输出峰值是k_d伏特，同时从公式(8)得到k_d也是鉴相器曲线在正交点附近的斜率(V/弧度)。因此，对于小的相位误差，输出近似是相位差的线性函数，

$$\langle v_{out} \rangle = k_d \phi \tag{9}$$

不仅希望输出对相位是线性的，也希望输出与幅度噪声无关。幸运的是，当正交关系得到满足时，这个灵敏度就自动地最小化，这可以从公式(8)看到。维持这个正交关系因此对这种类型的相位噪声测量因为几个原因就变得很重要。

在图 18.6 中的方框图显示锁相环如何保证所希望的正交关系。这个在虚线框中显示的锁相环被设计成有一个很窄的带宽，因为它的工作仅是在两个振荡器间建立一个标称值为90°的相位关系。因为锁相环在自身的带宽内能够使压控振荡器跟踪被测器件的输出，在小于 PLL 带宽的偏移频率的相位噪声就会被人为衰减，因此只有在偏移频率远大于锁相环带宽的相位噪声的测量是可信的。当设计锁相环内的低通滤波器(#1)时，这个限制一定要考虑进去。低通滤波器(#2)在其带宽范围内平均了乘法器的输出相位误差，它因此设定了这个相位噪声测量的分辨率带宽。

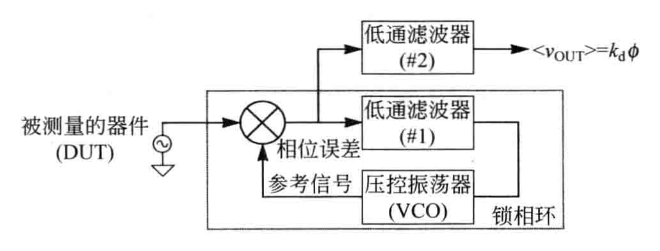


图 18.6 采用 PLL 的相位噪声测量

为了校准这个结构，要首先关闭锁相环，然后调整压控振荡器的频率以接近于被测装置(DUT)的频率。第二个滤波器通过了差频成分而阻拦了和频成分。用一个频谱分析仪或者示波器来监测这个滤波器的输出。频谱分析仪测量的是均方根值，因此要将这个值乘上$\sqrt{2}$以得到k_d。在得到了这个常量之后，启动锁相环。这个装置现在就能测量相位噪声了。

在实践中，在混频器的两个输入端口处几乎总是有两个隔离放大器(图中没有示出)，来阻止一个振荡器到另一个振荡器的不希望发生的锁相。这些振荡器应该有独立的电源供电，以防止通过电源噪声发生的注入锁定。如果一个寄生的锁相环被允许来控制整个系统，那么两个振荡器就会在寄生回路的带宽范围内彼此跟踪，导致在一些不被控制的带宽范围内不恰当地削减了测量到的相位噪声。

这个鉴相器方法通常要求压控振荡器具有比被测振荡器的相位噪声低得多的相位噪声，因为这里压控振荡器被用来作参考频率振荡器，它设立“正确”的过零点，来供 DUT 的输出进行比较。有可能先有一个校准步骤，来对参考振荡器的相位噪声进行表征，使得它能够从随后的测量中减去。这样做仅有的要求是参考振荡器的噪声足够小以允许实现可靠的减法运算。通常可以接受的情况是压控振荡器的相位噪声应该至少要像被测装置的噪声那么小，才可以进行精确的减法。

18.3.2 延迟线分辨器技术

尽管基于锁相环的测量可以有很精细的灵敏度，其缺点是需要两个振荡器，而且带来了伴随构造任何一个锁相环都不可避免的复杂度。仅需要一个振荡器的方法则是延迟线技术。尽管其精度比不上采用两个振荡器的锁相环技术，但是对于某些应用而言，这个方法的简单性使得延迟线技术更受欢迎。

这个测量的基础简单的就是一根延迟线的相位关于频率的关系。因为一个单纯的时间延迟表现出一个随频率增加而线性增长的相位偏离的负值，因而一个时间延时单元就可被看作是频率到相位的转换器，如图 18.7 所示。这个特定的频率（Hz 单位）和相位（rad 单位）的关系如下

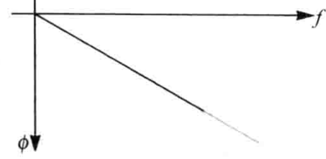


图 18.7 纯时延的相移

$$\phi = -2\pi f\tau \tag{10}$$

如果一个鉴相器被用来比较信号和它经延时后的信号，鉴相器的输出信号就对应于这两个信号之间的相对频率的波动。如果噪声的确是随机的，那么这两个输入信号的噪声就没有关联，此时一个准确的结果就能够得到。但如果两个信号的噪声是相关的，那么测量就会有错误。例如，$1/f$ 噪声引起了一个很慢的频率调制。如果延迟线的延迟相比于这些 $1/f$ 现象的周期要来得短，这两个鉴相器的输入就会同时移动，导致一个错的低相位噪声的结果。

因为频率的上下波动是通过式(6)与相位的波动相关的，这个技术引入了一个间接测量相位噪声的方法。输出从根本上说是与频率噪声成比例，如下所示：

$$\Delta\nu_{\mathrm{OUT}} = k_{\mathrm{d}}\Delta\phi = k_{\mathrm{d}}2\pi\Delta f\tau \Rightarrow \Delta f = \frac{\Delta\nu_{\mathrm{OUT}}}{k_{\mathrm{d}}2\pi\tau} \tag{11}$$

因此频率噪声的功率谱密度是

$$S_f(\Delta f) \stackrel{\mathrm{def}}{=} \frac{\overline{\Delta f^2}}{B} = \frac{\overline{\left(\dfrac{\Delta\nu_{\mathrm{OUT}}}{k_{\mathrm{d}}2\pi\tau}\right)^2}}{B} \tag{12}$$

反过来说，公式(6)将相位噪声和频率噪声的功率谱密度(PSD)互相联系起来：

$$S_\phi(\Delta f) \stackrel{\mathrm{def}}{=} \frac{1}{f^2}\frac{\overline{\Delta f^2}}{B} = \frac{1}{\Delta f^2}\frac{\overline{\left(\dfrac{\Delta\nu_{\mathrm{OUT}}}{k_{\mathrm{d}}2\pi\tau}\right)^2}}{B} \tag{13}$$

这也让我们对于小的相位误差将手写 L 近似为

$$\mathcal{L}(\Delta f) \approx \frac{S_\phi(\Delta f)}{2} = \frac{1}{2\Delta f^2}\frac{\overline{\left(\dfrac{\Delta\nu_{\mathrm{OUT}}}{k_{\mathrm{d}}2\pi\tau}\right)^2}}{B} = \frac{1}{2\,\Delta f^2}\frac{\overline{(\Delta\nu_{\mathrm{OUT}})^2}}{(k_{\mathrm{d}}2\pi\tau)^2 B} \tag{14}$$

这个测量装置因此如图 18.8 所示（以一个简化了的形式）来进行测量，采用一个频谱分析仪观测输出噪声作为频率的函数。因为频谱分析仪通常有差的噪声系数（例如，30dB），在分析仪之前加一个高增益低噪声放大器可以改进灵敏度。因为载波已被有效地去除了，在这个情况下采用一个前置放大器不会造成过载现象。

从公式(14)可见，测量的噪声功率必须用几个常数因子来乘，同时要用频率的平方Δf^2来相除，这是因为显示的实际上是频率噪声。作为一个数值的例子，假设 B 为 100Hz(20dBHz)，

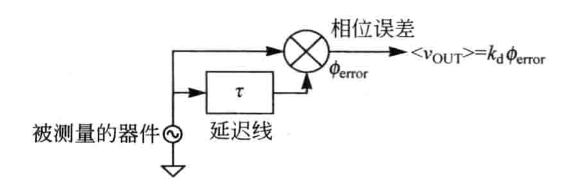


图 18.8 采用延迟线鉴频器的相位噪声测量(经简化了的)

$k_d 2\pi\tau$ 是 10^{-5} s。进一步假设频谱分析仪在 100kHz 处分辨率带宽内测得的功率为 2nW(−57dBm)。那么相位噪声在偏离载波 100kHz 处为 −110dBc/Hz。

一个微妙的地方是延迟时间一定要被选来产生一个在两个输入信号之间的正交关系。对于给定的振荡器的频率,这个要求隐指着允许的时间延迟的值是离散化的,这些值为振荡的四分之一周期频率的奇数倍。在实际情况中,或者一个变化的相移器与延迟线串联,或是振荡器的额定频率是可调的,这样就可以产生所要求的正交关系。如果在测量过程中,时间延迟或者振荡器频率发生了漂移,因为这个正交关系的不满足而产生误差。这个不希望的敏感性是区别于锁相环方法的明显的劣势。

如前面所提到的,一个较大的时间延迟允许对在载波附近的相位噪声更准确的测量。从公式(14),我们看到对于一个给定大小的相位噪声这样做也提升了输出电压值,因此就改善了在所有偏差频率处的测量灵敏度。对于延迟的长度上主要的限制在于需要将相位差保持在鉴相器的线性范围内。任意地假设鉴相器的线性范围为大约 1rad,那么延时的选择就一定要满足

$$\Delta\phi = 2\pi f\tau \leqslant 1 \Rightarrow \tau \leqslant \frac{1}{2\pi f} \tag{15}$$

这里 f 是最低的偏离频率,在该频率处相位噪声的测量也要保持合理的精确度。

作为一个典型的数值方法,假设希望研究在偏移频率为 1kHz 时表征相位噪声。公式(15)隐含着对于一个精确的测量时延必须在 160μs 数量级。要使这样大时延成为可能,需要大约 30km 长的典型同轴电缆!显然,这是一个不切实际的长度。即使有人愿意付钱来买这么长的电缆,在载波频率下的衰减将会使任何剩下的能进行测量希望打消。如果相当任意地选 10m 长的电缆作为上限(这依然是十分长的电缆),那么在偏移频率为几兆赫兹将代表一个实际的下限。对于低于这样的偏移频率的频率,灵敏度迅速变坏。

对于远处的噪声,这个用时延区分的方法在灵敏度上可以接近于锁相环的方法,但是在可以允许的时延上的限制使得如果采用物理手段实现的线,在接近载频的高精度相位噪声的测量成为不可能。如果愿意牺牲灵敏度,在偏移多少小于 1MHz(或许是接近 100kHz)处进行相位噪声测量是可能的。但是偏离载波的量再要些,准确的测量只能通过锁相环方法进行。

18.3.3 谐振鉴别技术

为了克服用长的电缆线实现足够时延所遇到的实际困难,工程师们发明了其他的替代方法。为了理解这些其他方法是如何工作的,首先回顾一下延时是正比于一个系统里相移对于频率变化的斜率:

$$t_D = -\frac{d\phi}{d\omega} \tag{16}$$

那么,不是采用一段电缆,可以用另一个其相移近似地随频率增加线性下降的元件来取代。正如其名字所表示的,一个通常的谐振器的相位与频率的特性可以被用在谐振鉴别方法中,见图 18.9。

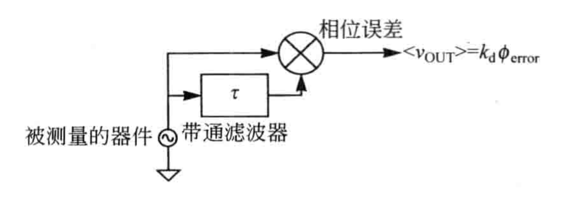


图 18.9　采用谐振器鉴频器的相位噪声测量装置(简化了的版本)

用了谐振器,存在着 Q 和相移之间的一个关系,亦即是 Q 与时延间的关系。为了决定这些关系,首先认识到一个带通滤波器(比如一个简单的 RLC 谐振槽路)的相移可以表示为如下形式

$$\phi = \frac{\pi}{4} - \arctan\frac{\omega/\omega_n}{Q[1-(\omega/\omega_n)^2]} \tag{17}$$

这里 atan 是反正切的简写,ω_n是中心频率。采用公式(16)对ω_n处的时延求值,最终得到

$$t_D\big|_{\omega=\omega_n} = \frac{2Q}{\omega_n} = \frac{2}{B} \tag{18}$$

这里 B 为-3dB 带宽,单位为 rad/s。这样,一个长的电缆的效应在某种程度上能够被一个有着恰当的低带宽的谐振器来模仿。这里的主要约束,正如所有的具有有限个极点的网络,总的相移量是有界限的,因此谐振器只能在一个有限的带宽范围内近似时延。公式(18)告诉我们,时延越大,带宽要越窄。在靠近通频带的边界处,时延会逐渐消失直到最后变为 0。延时接近于公式(18)给出的值的频率区间传统上(也是相当带任意性的)被取作是带宽的一半。

考虑标称延时区间外的系统的表现也十分重要。对于频谱中离开载波频率大于谐振器带宽的那些频率成分,滤波器的衰减会很严重。在这样的情况下,可以假设这些频谱成分不再存在,这样导致可以认为对鉴相器的一个输入是有效的不带噪声的。这样测量条件就趋近于基于锁相环的测量装置,而且输出就是相位噪声:

$$\Delta v_{OUT} = k_d \Delta\phi \tag{19}$$

总结一下,谐振器技术要求我们来测量在谐振器的半带宽范围内的频率噪声。这个频率噪声一定要用频率偏移量来按比例缩放以计算得到相位噪声。在谐振器带宽范围之外很远处,相位噪声能够被直接测量,不需要任何的按比例缩放。在这两个范围之间,显示的值是相位和频率噪声的某种混合,而且不要指望能够有什么有意义的测量。

重要的是要记住谐振器方法与延迟线技术同样要求标称的正交关系成立。因此需要一个可变的相移器(或有能力来变化振荡器的频率)来保证满足这个要求。

这两种鉴别器技术,尽管有局限性,因为它们的简单性都是可供业余爱好者实现的好方法。除非有非有不可的对很接近载波频率处振荡器相位噪声表征的需求,这些方法能够采用要求不高的仪器得到很好的结果。

18.4 误差来源

采用以上的任一种方法来进行精确的相位噪声测量，有必要来确定与量化各种误差的来源。可以想象得到，并不是所有的方法对于不同的误差的来源都具有同样的敏感性。然而，得到一个统一的噪声的模型还是可能的。在下面，噪声可以被假设为正好在频谱分析仪之前出现的可相加的噪声源。在这些技术之间的差别仅在于这个噪声源的实际测量结果。

18.4.1 频谱分析仪方法

为了看出哪些因素影响到频谱分析仪技术，回顾一下

$$\mathcal{L}(\Delta f) \approx \frac{S_\phi(f)}{2} \tag{20}$$

这也可以表述为

$$\mathcal{L}(\Delta f) \approx \frac{S_\phi(f)}{2} = \frac{\overline{\Delta v_{\mathrm{OUT}}^2}}{4v_{\mathrm{peak}}^2} \tag{21}$$

这里v_{peak}是混频器的最高输出电压。这个测量的噪声底限因此是被 Δv_{out} 最小值设定的，而 Δv_{out} 又是被混频器的噪声和频谱分析仪的有效输入噪声所决定的。

$$\mathcal{L}(\Delta f) \approx \frac{S_\phi(f)}{2} = \frac{\overline{\Delta v_{\mathrm{OUT}}^2}}{4v_{\mathrm{peak}}^2} = \frac{1}{4}\,\frac{\overline{\Delta v_{\mathrm{n,mixer}}^2} + \overline{\Delta v_{\mathrm{n,SA}}^2}}{v_{\mathrm{peak}}^2} \tag{22}$$

混频器的输出噪声和频谱分析仪输入噪声被假设为不相关的，因此它们的功率可以被直接相加，如公式(22)所示。很明显，灵敏度敏感能够通过采用能够输出大的输出且噪声系数小的混频器来得到改善。通常采用二极管环形混频器，其中每一支路是几个串联起来的二极管。在频谱分析仪之前加入低噪声放大器特别有帮助，因为低噪声放大器能够有效地减低输入噪声，$\Delta v_{\mathrm{n,SA}}$。商用的仪器能够取得-180dBc/Hz 那样低的噪声底限。

18.4.2 延迟线鉴别器

在前面讨论的两种鉴别器方法中，频谱分析仪都直接显示频率噪声，因此要得到相位噪声，使用者一定要用频率偏移量的平方来将结果归一化。因此，任何在频谱分析仪输入端出现的噪声表现为有效的频率噪声：

$$S_f(\Delta f) \overset{\mathrm{def}}{=} \frac{\overline{\Delta f^2}}{B} = \frac{\overline{\left(\dfrac{\Delta v_{\mathrm{OUT}}}{k_{\mathrm{d}}2\pi\tau}\right)^2}}{B} = \frac{\overline{\Delta v_{\mathrm{n,mixer}}^2} + \overline{\Delta v_{\mathrm{n,SA}}^2}}{B(k_{\mathrm{d}}2\pi\tau)^2} \tag{23}$$

在很多实际的实现中，一个缓冲的放大器被插入在延迟线输出和混频器输入之间。如果采用这样的一个放大器，其噪声贡献也一定要被考虑。因为这个放大器输出的噪声出现在混频器的一端，它也表现为频率噪声项，尽管要用一个等于鉴相器增益的因子进行按比例缩放。因此，总的频率噪声功率谱密度由下式给出：

$$S_f(\Delta f) \overset{\mathrm{def}}{=} \frac{\overline{\Delta f^2}}{B} = \frac{\overline{\left(\dfrac{\Delta v_{\mathrm{OUT}}}{k_{\mathrm{d}}2\pi\tau}\right)^2}}{B} = \frac{\overline{\Delta v_{\mathrm{n,mixer}}^2} + \overline{\Delta v_{\mathrm{n,SA}}^2}}{B(k_{\mathrm{d}}2\pi\tau)^2} + \frac{\overline{\Delta v_{\mathrm{nm,buffer}}^2}}{B(2\pi\tau)^2} \tag{24}$$

相位噪声测量的底限因此为

$$\mathcal{L}(\Delta f) \approx \frac{S_f(\Delta f)}{2(\Delta f)^2} = \frac{1}{2}\frac{1}{B[2\pi(\Delta f)\tau]^2}\left(\frac{\overline{\Delta v_{\mathrm{n,mixer}}^2} + \overline{\Delta v_{\mathrm{n,SA}}^2}}{k_{\mathrm{d}}} + \overline{\Delta v_{\mathrm{n,buffer}}^2}\right) \tag{25}$$

注意，这个值在低频时增长很快，这与我们早期对这个方法弱点的观测一致。

18.5 参考文献

下面的来自于惠普(Hewlett-Packard)的笔记包含有很多有关相位噪声测量的有用信息："Phase noise characterization of Microwave Oscillators"(微波振荡器的相位噪声表征)(产品说明 11729B-1，1983 年 8 月)和"The Art of Phase Noise Measurement"(相位噪声测量的艺术)(Dieter Scherer，RF & Microwave Measurement Symposium and Exhibition，1984 年 10 月)

第19章 采样示波器、频谱分析仪与探针

19.1 引言

示波器和频谱分析仪是任何射频实验室中都具备的测试设备。原因很简单：同时从时域和频域上分析信号是很有用的，尽管这两种表象从理论上来说提供的信息是等同的。

大多数电气工程师对较低频的示波器的基本操作原理都是相当熟悉的。但是仍有相当一部分人对探针是如何工作的理解(特别是有关接地技术)不完整。随着频率的增加，这种片面理解所造成的影响也就越严重。在对传统的低频示波器进行简要的回顾之后，我们的主要关注点是吉赫兹(GHz)频率下使用示波器时所需额外考虑的一些问题。同时由于在高频条件下通常使用的采样示波器会有一些不被注意的模式，它们会诱发所谓的“飞行员错误”，因此我们将花费一些时间来研究一下它们是如何工作的，及如何避免被这些工作模式所迷惑。高速采样电路本身就是有其独特之处，因此这种示波器也给我们一个理由来花一些时间了解取样器是如何工作的。

另外一个令人惊叹的仪器是现代的频谱分析仪(其价格也大约与“惊叹程度”的平方成正比，译者注：开玩笑的话)，它能够在一个大的动态范围(比如 80～100dB 赝空动态范围，SFDR)以及很宽的频域跨度(例如在单一的一个仪器中实现 DC 到 20GHz)进行测量。

要想最大限度地使用这个仪器，并且避免常见的测量错误，去理解它们的内部结构以及探针或者其他连接测量仪器与被测器件的插件的特性是十分重要的。下面先对通常的时域连续示波器进行一个简要的回顾。

19.2 示波器

通常可以有把握地假设大多数工程师对一个典型模拟示波器是很熟悉的，但是现代的实验室和教室中越来越多地用到数字化仪器。学生和工程师因此越来越少遇到普通的示波器。而如今更典型的情况是，一层层的软件将用户与自然世界隔离开来了，而阻碍了人们对深层次的重要理解。所以，我们有必要花一些时间来快速地回顾一下经典模拟示波器。

19.2.1 “纯”模拟示波器

图 19.1 给出的是一个典型的连续时域(即非采样)示波器。精密的宽带衰减器接收输入信号，并用一个经校准的量来对其进行按比例缩小，以避免垂直放大器的过载。垂直放大器模块能提供足够的增益来驱动阴极射线管(CRT)的偏转板。垂直放大器后接的一根延时线给触发器和扫描电路提供足够的时间，以在输入信号到达 CRT 之前就开始工作。这个延时因而允许示波器显示波形的那些部分，其实际上比触发过程还要早到达。延时线可以用一条专门设计的低失散度电缆，或者是一条传输线的集总元件近似。如果是前者的话，

那么线长可能要有好几米。无论是前者还是后者，延时元件的衰减和失散特性对示波器的工作都是至关重要的。均衡电路是补偿延时线失真无一例外被使用的方法。

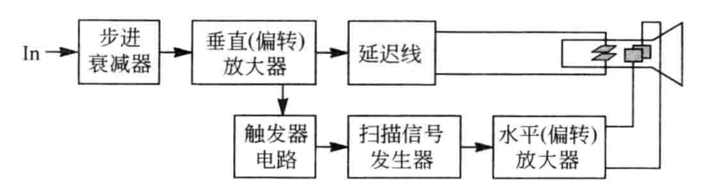


图 19.1　典型的模拟示波器的方框图

所有的经典模拟示波器使用的都是静电偏转示波管，因为磁偏转线圈的电感太大，无法允许对比几兆赫兹频率高的信号进行操作。电视和电脑的显示器使用磁偏转，是因为通过使用更多的圈数，可以很容易地实现大的偏转灵敏度，从而显著地减小了给定屏幕尺寸下示波管的深度(所以重量也减轻了)。要求一般的偏转带宽使得磁偏转在这些应用中得以采用。静电偏转则是高频示波器唯一可行的选择，其代价是要有相当长(深)的显示管。

扫描发生器产生锯齿状的电压来驱动水平偏转板。锯齿电压使得示波管内的电子束线性偏转，从而产生恒定速率的水平扫描。一个触发电路通过输入信号来使这个锯齿波同步，其过程是只有在输入信号显现出一些预定的特征，例如超出一个特定的电压阈值时，才启动扫描，从而使得锯齿电压和输入信号在时间上同步。为了在没有信号时也能产生一个令人满意的基线，示波器设置了可选的触发模式(“auto”)，在此模式下如果一段时间内没有触发事件发生，锯齿电压波便会自动启动。

这种基本的示波器结构问世已经有相当一段时间了。1897 年，晶体整流器的发明者费迪南德·布朗(Ferdinand Braun)发明了示波器的一个原始形式①。因此，示波管在德国依然有时被称作是“布朗管(Braunsche Röhre，即 Braun Tube)”。在随后的几十年中发生的改进包括大幅地提高了带宽和灵敏度。在 20 世纪 30 年代，静电偏转的示波管已经进步到可以使得几十伏特的电压能产生电子束的 1cm 偏转。这种示波管至今仍在使用并且可以与直至 100MHz 带宽范围兼容。

当频率进一步增加时，从实际的放大器中获取大的电压摆幅就变得越发困难了。来为了认识这个困难程度，我们来看一下一个振幅 100V，1GHz 的正弦波，最大斜率超过 600V/ns，或者说 600kV/μs。要构造一个能够提供 100V 峰值电压并且能同时产生如此高摆率的放大器绝对不是一个简单的任务！

为了能理解提高偏转灵敏度的发展过程，要注意到普通的偏转板只在很短的一段时间内有机会使电子束发生偏转。为了解决这个问题，利用分布式垂直偏转结构来替代简单的平行板结构，以提高电子束受到偏转电压作用的时间长度。在这样一种结构中，每一块偏转板都被分成很多块，并通过一个小电感将临近的块连在一起。因而这种分布式偏转“板”实际上就是一条传输线(如图 19.2 所示)。实际上，整个分布结构可以看作是一个由螺旋缠绕

---

① “Ueber einVerfahen zur Demonstration und zum Studium des zeitlichenVerlaufes variable Strome” [On a method of demonstrating and studying the time dependence of variable currents](一种演示与研究时变电流的时间依赖关系的方法)，*Annalen der Physik und Chemie*，*February* 1897。这个第一个示波器是由磁场来进行偏转的，是用来研究电力设备的输出。当时，并不存在电子放大器，因此这个早期的仪器灵敏度是相当低的。

的单元块组成的匣子。这种结构安排使得电容和电感连续地分布在整条线上从而使得传输线的有效带宽最大化。

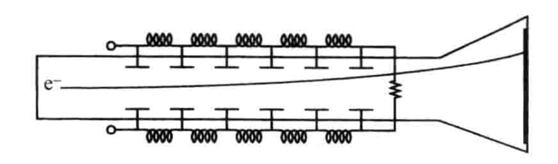


图 19.2 分布式的垂直偏转阳极板

为了使电子束受偏转电压作用的时间尽量长，电子的运动速率与传输线的延时是相吻合的。垂直偏转灵敏度可以被提高一个数量级，达到约 3V/cm，从而允许示波器工作在超过 500MHz 的频率下，且具有经放大后 10mV/cm 的灵敏度[②]。对摆幅的要求因而大为下降至 30kV/μs(但这依然是一个很大的数目)。

将这个通用结构扩展到吉赫兹范畴是在 1970 年末成功实现的。这个带宽的改善主要是因为 CRT 设计的进步而实现的[③]。采用具有水平和垂直偏转板的分布结构以及微通道板的示波管可以使偏转灵敏度再提高三倍，这样就大大地放宽了对放大器设计的要求。

微通道板示波器管利用了这样一个简单的观察：在电子束获得足以在屏幕上产生一条亮迹的高能量之前对其进行偏转要更为容易一些。所以，我们就这么做做看。为了在偏转后获得足够的电子束能量，低能量的电子流流过一个充满了微小柱形通道的薄半导体板并在通道的两端加有足够高电压以产生很大的加速电场(见图 19.3)。当电子撞击微型通道壁时，它们会释放出次级电子，使得电子束电流成倍提高，从而在荧幕上产生一条很容易看见的亮迹。这种示波管极高的灵敏度，使得 Tektronix 7104 让单个 350ps 长的事件在普通室内光条件下也能被肉眼观察到。同时，它还支持超过光速的扫描速度！(这又是作者的调侃，译者注)

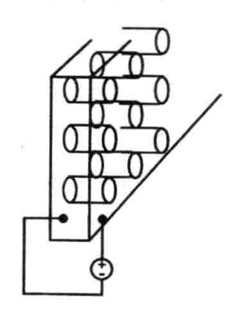


图 19.3 微通道板

这种结构的优势在于：在噪声、失真以及测量夹具的质量限定下，观察者所看到的就是实际发生的。不幸的是，这种高保真度需要以异常复杂的示波管技术为代价。更糟的是，我们还不清楚如何将这项技术扩展到远大于 1GHz 的带宽范围。这个扩展所带来的挑战是极其大的：更快的扫描速度就意味着电子束在荧幕上停留更短的时间，使得荧幕上的亮迹变暗。为了提高亮度而提升加速电压则会使得电子束更难偏转，即便使用的是微通道板。而且支持高频操作的晶体管为了支持高带宽必须要有更小的特征尺寸，因而击穿电压就更低，

② 能做到这一点的第一批商用产品，即 Tektronix 7904 的 CRT 只是垂直偏转板是分布式的，水平偏转电极依然是普通的平行板。如果其他的“经典”的示波器，这些示波器可以在旧货市场以相当不错的价钱买到，它们可以为业余爱好者认真考虑作为一个选择。而且标称的 500MHz 并不是限定死的，因为用户可以设法直接驱动垂直的偏转板。最后得到的灵敏度是大约 3～4V/cm，同时带宽可以大大超过 1GHz。

③ 实际上，GHz 示波器在大约 1960 年就首次成为商用产品，但是这些示波器将输入信号直接连到 CRT 的偏转板上。灵敏度因此是差的(要得到任何分辨得出的偏转就需要好几伏)，结果是这些仪器只有有限的用途。偶尔，也可以在旧货市场中看到这些仪器，但是最好不要去买(它们是如此之笨重，除非你想去扭转重力场：译者注)。如果与型号为 7A29 的可插入放大器一起用，7104 在灵敏度为 10mV/cm 时，有一个标称值为 1GHz 的带宽，而且如果 CRT 的偏转板是直接被驱动的(此时，大约 1V/cm 的灵敏度依然可以得到)，它们可以一直工作到 2～3GHz。

进而降低了摆动幅度。类似的挑战还有不少。

由于这个技术的局限(称为"游戏规则")，工程师们理所当然地选择了另外一种"游戏"方式。

19.2.2 采样示波器

这另外一种"游戏"方式指的是采样。寻求一个采样结构将设计难题转移到了一个纯粹的电子学领域，使得可以用只比传统电视中所用稍微复杂一点的 CRT。现代采样示波器结构的第一篇文章出现在 1950 年④。

作为采样示波器基础的关键思想可以用图 19.4 得到最好的说明。在这幅图中，高频正弦波在黑点标明的瞬间被采样。被采样的波形然后经滤波来产生图中所示的低频正弦波。这个低频正弦波在传统的低频示波器中很容易显示出来。通过这样的描述，我们可以清楚地看出采样示波器类似于一个闪光灯，使得波看起来变慢。这样来想象采样示波器，对理解它的性质非常有用，也包括了解它可能引起的误导方面。

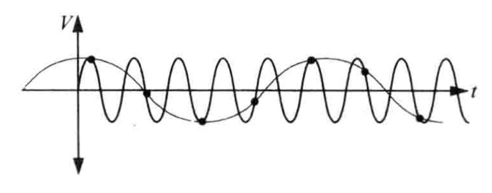


图 19.4　取样示意图

显然，这个结构的奥妙之处就在于采样电路本身。除开触发和扫描子电路的精妙设计，剩余部分则是相对普通的电路。高频的限制是被采样电路的时间分辨率所控制的。采样窗口(即孔径)越窄，带宽就越高。近似地说，窗口宽度和带宽(单位 Hz)的乘积大概是 0.35。几皮秒的采样孔径代表了商用仪器能达到的最高性能，这对应着超过大约 50GHz 的带宽。基于超导技术制成的采样器则能够在更高的频率条件下工作。实验室条件下已经证明了在几百 GHz 下操作的可行性。

鉴于采样门电路是此类示波器的核心部分，我们有必要来仔细看一下实现如此窄的采样孔径所用的一些方法。图 19.5 的方框图给出了 20 世纪 60 年代早期到 80 年代所使用的典型采样器的结构⑤。在这里，晶体管的作用是作为一个雪崩脉冲发生器，类似我们在第 8 章里面所描述的那样。集电极电源电压被调整到略低于可以引起自发雪崩的电压值。当基极收到一个正触发脉冲时，晶体管就会达到完全雪崩击穿的状态⑥。于是集电极电压骤降，而在此同时发射极的电压在此同时暴增。这里典型的过渡时间大概在几百皮秒左右。由于自身特点，这样雪崩发生器产生的信号都是极其得快，并且能够直接运用到采样桥操作中而达到带宽直至 1GHz。

为了达到更快的工作，雪崩发生器的输出需要调节至可以产生更短的上升下降时间。一个起到共模扼流圈作用的宽带变压器首先移除任何由集电极和发射极延时的略微差异产

---

④ J. M. L. Janssen，"An experimental 'stroboscopic' oscilloscope for freqnencies up to about 50Mc/s - i: fundamentals"(频率高达 50Mc/s 的一个实验的"采样"示波器，第一部分：原理)，*Philips Tech. Rev.*，v. 12，no. 3，August 1950，pp. 52-59。实际的硬件用到一个五极真空管来作为采样元件，可以在直到 30MHz 频率处工作。

⑤ 这个示意图是采自于 Tektronix S-6 14GHz 型号的采样头电路。

⑥ 在取样示波器中利用雪崩现象很明显是属于 G. B. B. Chaplin *et al.*，"A sensitive transistor oscillator for sampling oscillograph with D. C. to 300 Mc/S bandwidth，"(一个具有从直流到 300 Mc/s 带宽的灵敏晶体管示波器)，*Int'l Transistors and Association of Semiconductor Developers Convention*，May 1959，pp. 815-823。同时见 U. S. Patent(美国专利)#3,069,559，1962 授权。第一个商用示波器是在 1959 年由 Lumatron 提供的，尽管 HP 185A 以及其 500MHz 取样插入单元在 1960 年的推出在这个领域中其后的发展有更大的影响。

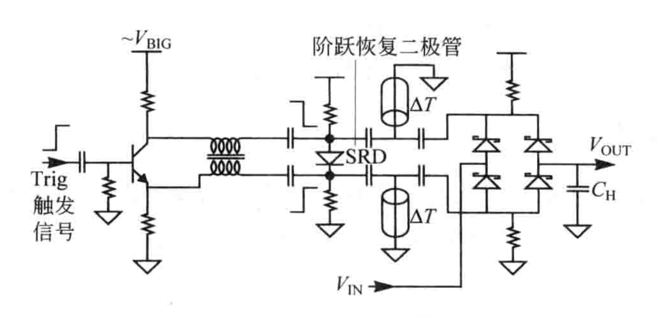


图 19.5 具有代表性的采样门电路(简化版本)

生的小的共模成分⑦。变压器输出端的这个近乎完美的差分电压使得通常是正向偏置的阶跃恢复二极(SRD)反向偏置。最终二极管瞬间关闭，生成一个少见的极快的阶跃(正如第 9 章所描述的)。SRD 的特性是其上升时间可以被加速一个数量级，达到 10ps 范围⑧。

通过使用末端短路的传输线段，这个快速阶跃被转化成一个窄脉冲。如果每个线段都有一个单向的飞行时间(time-of-flight)延时 ΔT，那么总的脉冲宽度就是在 $2\Delta T$ 数量级。对应于 10ps 的典型上升时间因此就是大概 20～30ps 的脉冲宽度。常见的传输线段的典型延时大概为 50ps/cm，所以快速采样器需要使用短的传输线(一般来说有几毫米长)。低的色散对这个结构来说十分重要，因此传输线常常使用空气作为介质。

这些窄脉冲被用来驱动，比如说，一个由四个二极管构成的开关，这些二极管在通常情况下处于截止状态。这些脉冲使二极管正向偏置，从而产生一个从输入到输出的导通路径，使得输出保持电容C_H被充电到与输入相同的电压值。这个采样-保持(sampled-and-held)电压只在采样重复频率(可能只有几千赫兹那么低)改变，所以后续的阶段就不需要特别高的带宽了。事实上，图 19.5 所示的电路可以运用在一个标准示波器的前面来将其带宽扩展到远比吉赫兹高的范围(尽管生成一个合适的触发器依然是一个挑战)。

将雪崩发生晶体管替换为 SRD 可提供大于一个数量级的加速，从而允许采样示波器在直至 20GHz 的信号下仍能正常工作。另外一个类似的提高因素可以由冲击波(shockwave)传输线提供⑨。想象一条人为的传输线，其中分布电容由半导体结的依赖于电压的电容提供。如果我们想象一下下降的电压使得反向偏置结电压上升，那么那些边沿就将在其下降时经历减小的电容，因而得到了加速。沿着这样的线上传送下去脉冲，将经历在下降沿提速的不断改善，直至受到其他机理的限制(例如色散或有限的传输线间的结的带宽)。阶跃上升时间低于～700fs 已经有报道，这相应于带宽超过～700GHz，其实现方式是通过使用沿着一条共面线连接的肖特基二极管⑩。

---

⑦ 在要求不那么高的应用中，可以不用这个变压器。而对相对相位的补偿简单地用调整到集电极与发射极的互连线长度来提供。

⑧ 在 1962 年推出的 HP 186A 是采用 SRD 的第一个仪器。

⑨ 冲击波传输线基于的基本思想早在 1960 年就被描述了，但直到 Mark Rodwell 在 20 世纪 80 年代后期发表了他的有关这类传输线的工作，这个想法停留在一个学术的好奇上。

⑩ U. Bhattacharya, S. T. Allen, and M. J. W. Rodwell, “DC-725GHz sampling circuits and sub-picosecond nonlinear transmission lines using elevated coplanar waveguide”(直流到 725GHz 的采样电路与采用抬高的共面波导亚皮秒非线性传输线), *IEEE Microwave and Guided Wave Lett.*, v. 5, no. 2, February 1995, pp. 50-52。抬高了传输线可以减少能量与半导体衬底的耦合，从而减少了损耗及色散。

19.2.3　探针、补偿、噪声与接地

注意到下面这一点是十分令人吃惊的，即有那么多工程师们为了有灵敏的频谱分析器以及高速示波器付出了相当多的努力——但却在与其相配的探针使用上导致高噪声、低带宽、不平的通频带，以及其他的错误的测量。为了理解这个问题，需要先来考虑一下示波器探头的详细特性。

与我们可能想到的恰恰相反，一个探头绝对不只是一段一边是尖端，而另一边连着连接器的线而已。考虑这样一个事实：大多数的 10:1 探头(这么命名是因为它能够提供 10 倍的衰减)在测试条件下对被测的电路呈现的阻抗是 10MΩ，而同时可以提供上百 MHz 的带宽。但是直觉告诉我们，符合这样的带宽条件所允许的最大电容应该低于约 100aF(0.1fF)！所以，探针如何在提供 10MΩ 阻抗同时提供这么大的带宽呢？问题的答案是探头和示波器的组合并不是一个“常规”的 RC 网络。

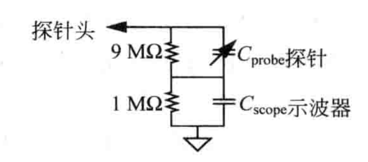


图 19.6　以示波器为负载的通用探针的简化模型

图 19.6 给出了一个简化的示波器-探头组合的简化模型。1MΩ 的电阻代表的是示波器的输入电阻，C_{scope}代表的是示波器的输入电容。

在探针内有一个 9MΩ 的电阻来提供低频下所要求的 10:1 衰减[⑪]。为了避免这个 9MΩ 电阻的使用所造成的极大的带宽减小，探针还有一个与电阻并联的电容。在高频条件下，10:1 的衰减实际是由电容分压器提供的。可以证明(而且并不难证明)：当顶部的 RC 与底部的 RC 相等时，衰减倍数恰恰是 10，并且这个衰减倍数不受频率的影响。存在着一个零点恰好精确地消除了慢极点的影响，从而使得传递函数没有带宽的限制。

由于要想用固定的元件来实现完美的补偿是不可能的，所有的 10：1 探针都设置了一个可调电容。为了更透彻地了解这种调节的必要性，我们用下面这个传递函数(参见第 12 章)来回顾一下不完全的零极点消除的效应：

$$H(s)=\frac{\alpha\tau s+1}{\tau s+1} \tag{1}$$

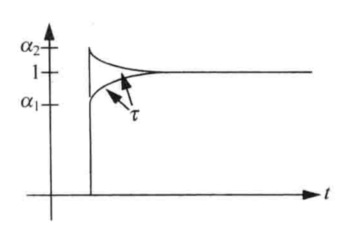


图 19.7　极点-零点对的可能的阶跃响应

α 的初始值就是 α，而最终值是 1。我们可以画出(见图 19.7)一个$\alpha_1<1$ 和一个$\alpha_2>1$ 的阶跃响应。在这两种情况下，响应都立即跳到 α，但是最终都会以极点的时间常数稳定到最终值(即 1，译者注)。如果 $\alpha=1$，那么响应到达最终值的时间为零，尽管在所有实际电路中都会有另外附加的极点，这使得上升时间被限制到一个非零值。

从图 19.7 中，我们可以很容易地看出调整电容来避免大的测量误差的重要性。这种补偿可以很简单地实现，即通过仔细观看对一个足够低频率的方波响应，使得按照零极点对的缓慢稳定的过程变得可见，然后调整电容以获得最平的时间响应。很多示波器专门为探头校准设置了一个方

⑪　用在极高频率下的探针通常采用带有分布式电阻的特殊的电缆。

波输出。在测量前需要做这样的校准，否则可能会造成较大的测量误差。

另外一个常见的错误是使用大多数探头提供的接地引线。尽管方便，长的接地引线几乎一定会导致错误的测量结果。原因很简单：地线的电感（典型的大概 0.5nH/mm）会增加一个阻抗，这个阻抗与示波器的阻抗会形成一个低通滤波器。在靠近截止频率的频率处，可能出现共振峰值。

我们来看一个例子（见图 19.8），假设接地线有 20cm 长（不幸的是，这并不是一个非典型的值）。因此寄生电感大概会在 50～100nH 左右（确切值受到路径的形状）。给定一个 15pF 的典型示波器输入电容，形成的低通滤波器的截止频率大概低于 200MHz。显然，这样的结构对于一个 1GHz（或更高频率）的示波器来说没有意义，带宽在信号到达示波器之前就已经被限制在低的值了！除了带宽限制，频率响应也与平坦相差甚远滑，这会造成错误的幅度的测定。

除了作为一个电感外，接地线也可以与示波器形成一个相当有效的环形天线，这将收集到很远很宽的范围内的噪声。常见的示波器亮点轨迹的模糊并不像很多人想象的那样是由于内部因素。将与地的连接长度减少到最小，可以很好地消除这种非正常现象。

为了避免这类问题，不要使用长的接地线。就是这样。赶紧将它好好收起来，放在你的工具箱里，让它攒灰吧。代之以，我们使用像弹簧一样的小线圈围绕在探针的接地环上。如果这些"弹簧"弄丢了的话（这种情况经常会发生），可以很容易地用一段合适长度的导线做出一个替代品，就像图 19.9 所示的那样。本书的作者往往喜欢将这些结构都焊在试验板的关键测试点上，以省去这之后再寻找这些地"弹簧"的麻烦。电感值减小了约两个量级，从而允许了探头以其所支持的最大频率进行测试。

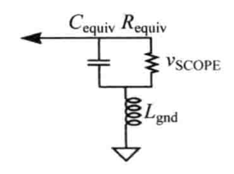


图 19.8　示波器与带有接地引线的探针电路模型



图 19.9　对于更高频探头的接地设置

商用的 10：1 无源探头普遍可以满足在高达几百 MHz 的条件下使用。在 GHz 频率下，大多数（但不是全部）探头都是 50Ω，以维持与其他微波设备的兼容性。遗憾的是，这么低的电阻值限制了那些电路节点能够被测试；高阻抗的节点会因为探头的存在而发生阻抗的大幅度下降，从而使得测量结果失去意义。为了解决这个问题，可以使用一个有源场效应管探针，这种探针在靠近探针处有一个源跟随器，使得阻性和容性的负载效应降至最低。这种电容的减小（其值可低至约为 100fF～1pF，如果加入补偿则会更小）容许有较高的节点阻抗（例如 1～100kΩ），这也是以 DC 稳定性的降低、绝对精度的降低以及带宽的减小为代价的。但是这种探头可以在超过几 GHz 的条件下工作，这对绝大多数情况都已经足够了。一个例子是 Agilent Model 1158A 10：1 的有源探头，在探头体内的集成缓冲器可使其工作带宽超过 4GHz，并且可以分别提供 100kΩ 的输入电阻和 800fF 的输入电容。

对于微波集成电路上（或者是非常小的分离模块）的节点的探测有着它自己的特殊问题，既是因为涉及到的机械上的挑战，也是因为要想避免从集成电路到探头的不良的过渡是

十分困难的，然后，最终是关系到探头与什么样的仪器相连接。这些困难确实很棘手，但幸运的是它们并非是无法解决的。例如，GGB Industiries Model 35有源探头的工作范围可以从直流到26GHz，具有1.25MΩ的输入电阻和50fF的输入电容。一个50Ω无源探头Model 110H可在高达110GHz的条件下操作，这个上限频率可与一个1.0mm W同轴连接器兼容。通过使用一个WR-5的波导连接，Model 220可将工作上限提高至220GHz[⑫]。当然，波导连接也意味着无法在直流条件下工作。事实上，如果想要维持在一个单一模式下操作，带宽就不能超过这个频率的一倍。

当使用这类探头在矢量网络分析仪（vector network analyzer，VNA）进行阻抗的测量时，有必要先进行校准，就像用任何其他VNA对阻抗进行测量一样。为了便于进行TRL或是LRM校准操作（见章节8.4.2）更加简便，一些公司提供了特殊的基片，在这种基片上各种校准结构已经被制造出来了。它的几何结构几乎总是共面的，并且经常采用地-信号或是地-信号-地的布局方式。

用于微波测量的探头，无论是有源的还是无源的，都是相当昂贵且容易被损坏。对学生和业余爱好者来说，一个经济的折中选择是基于基本的、如图19.10所示的10∶1示波器探头作一些小的修改来构造你自己的无源探头。电阻R被选择来提供需要的基本衰减，而补偿电容C则被选择来使得带宽最大，在这个带宽内，响应能在一定的偏差内足够平滑。表19.1给出了与不同衰减倍数对应的R与C的值。

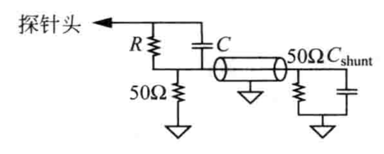


图19.10　N∶1手工制作的无源探针的基本原理图

表19.1　探头的元件值

衰减	R	$C_{并联}/C$
10∶1	225Ω	9
20∶1	475Ω	19
50∶1	1.225kΩ	49
100∶1	2.475kΩ	99

电容C_{shunt}代表的是连接在探头上的仪器的输入和杂散电容，以及跨在探头本身内的终端电阻两端上的寄生电容。给定了这样一个事实，即要能在大的带宽下工作的系统一般会被设计成近似有纯实数50Ω的输入阻抗，那么C_{shunt}的典型值就应该约为1pF（如果有那么多的话）。大致看一下表19.1所示的关键元件的典型值，我们就应该停下来再思考一下了，因为所需要的补偿电容随着衰减倍数的增加而急剧减小。例如，如果试图构造一个100∶1的探头来产生2.5kΩ的输入电阻，所需要的电容大概为10fF。利用手工方法是不可能有把握地生产出这么小值的电容的。在任何情况下，补偿电容都需要是可调的。一个10∶1探头的电容还比较容易实现，但是一个250Ω的输入电阻对于许多应用来说就太低了。较好的衰减选择应该是20∶1或50∶1，但是这些仍然需要提供亚皮法拉级的电容。

⑫　在一个WR-X波导的标识中的“X”指的是关键的（宽带）尺寸，单位是百分之一个英寸。这样，WR-5有一个50密尔（mil，即毫英寸）（不是毫米）的关键尺度，或者说1.27 mm。在截止频率处的波长大约是2.54 mm，这对应于大约120GHz。因此可以期待WR-5工作的频率范围应该比120～240GHz稍小，的确这个波导的性能指标上指出的频率范围是140～220GHz。

避免这些问题的一个方法是不去改动探针的电容 C_{comp}，然后通过增加 C_{shunt} 值来施加补偿。采用这些想法来构造一个探头，将所有以微带形式出现的元件组装在一块合适的电路板上(例如 FR4)。可以将一根小针与一个表面安装(surface-mount)电阻 R 串连焊上去作为探针。根据所选择的衰减倍数，所需的 C_{shunt} 可由商用微调电容器来提供(例如具有很低的串联电阻和电感的活塞式电容器)，或利用 1.6mm 长的 FR4 的 2.5pF/cm^2 电容。用铜箔片制造电容，并通过对箔片的增加或修剪来实现电容的调节。于是探头就大致如图 19.11 所示的那样。

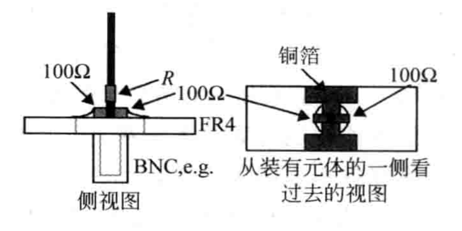


图 19.11　手工制作的探针

利用这种方法制造 20∶1 和 50∶1 的探头是完全实际可行的。通过对铜箔片的仔细修剪，很容易实现低吉赫范围内的带宽。给探头套上热收缩管还可以提高机械稳定性。

19.3　频谱分析仪

有几种电路结构被用来构造频谱分析仪，但是商用 RF 分析仪全都是基于超外差式的。见图 19.12。输入信号可能经过某种衰减后，与一个本地振荡器混频到中间频率(中频，IF)，这个本机振荡器的频率由一个扫描发生器(ramp，或是它的一个数字电路实现)控制。这个扫描信号同时提供给显示屏的水平偏转电路来建立一个频率轴。这个本地振荡器担负着完成起来困难的任务，要在非常宽的频率范围(例如，几十倍大小)下工作。本地振荡器的核心部分通常是一个 YIG(Yttrium-Iron-Garnet 钇铁石榴石)球。

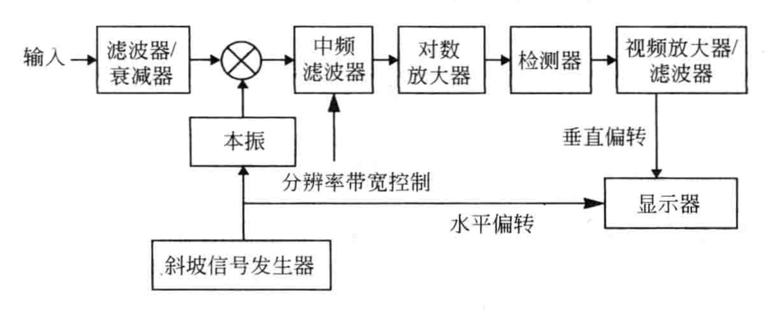


图 19.12　扫描频谱分析仪的方框图

这个器件(见图 19.13)是压电晶体的对仗，因此它的行为类似于一个带有串联寄生电感的并联 LC 共振器工作(这与作为石英晶体模型的两端口带有并联电容的串联谐振器相对仗)。典型的无载 Q 值超过10^4，从而具有很好的频谱纯度。与频谱分析仪有着更大联系的是，共振频率与外加磁场的强度是线性成比例的，从而允许在特别大的范围进行调谐，这个频率范围是实验室设备所需要有的。

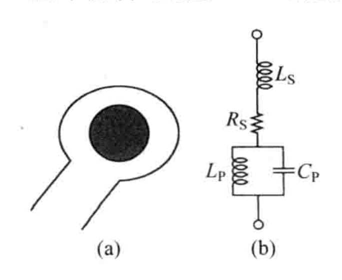
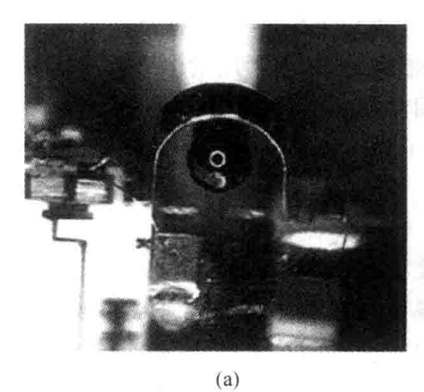


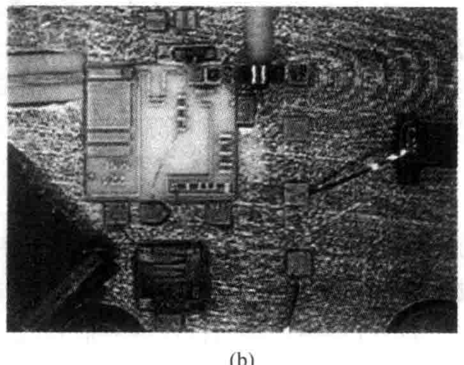
图 19.13　YIG 球面，耦合回路，及一个可能的等效电路

如图 19.14 所示，球体通常被安装在介质支持杆的末端。一个由宽的导体带制成的单圈耦合环将球围住，并与

振荡器其他部分连接起来。图中没有画出的是频率控制磁体以及相关联的电子电路。全部这些结构将会被封装在一个电磁屏蔽很好的结构内，以避免周边场对振荡器的干扰。



(a)



(b)

图 19.14　YIG 球面与耦合回路(左图)，以及被使用的实际环境
(得到 David Straight，M/A-Com 的许可)

中频(IF)滤波器的噪声带宽被称作“分辨度带宽”，它的值可以从几个离散的值中选。这个带宽决定了谱功率被测定的频率范围。IF 滤波器通常由级联的互相同步的调谐带通滤波器级实现，总的结果是一个高斯响应波形。这样的一个滤波器拥有很好的相位线性(或者说近于恒定的群延时)，其噪声带宽是－3dB 带宽的 1.2 倍。在积分噪声密度以得到总的功率时，需要考虑这个噪声带宽(依比例约为 0.8dB)的差别。

为了提供一个以分贝度量的幅度测量，一个对数放大器将 IF 滤波器的输出压缩。因为这样一来噪声峰值相较于平均值而言幅度被减小了，这个对数放大器会引入某些测量失真。在这种情况下，功率的测量值就要低大约 1.45dB。一个常见错误是没有将对数失真考虑进来。要记住，举例来说，不能简单地将所有分贝表示的测量结果取平均。必须首先将原始数值取平均值，然后再转化为分贝值。

经过滤波和对数放大处理后，一个“视频”检测仪会驱动显示屏的垂直偏转电路[13]。另外一个滤波器(被恰当地称为视频滤波器)用来使得探测仪的输出变得平滑。这个检测器在测量随机噪声上的特征引入了一个自身校准的问题。大多数检测仪是被校准来读出均方根值的，而这是基于正弦波的假设，即将原始的输出乘上 $\frac{1}{\sqrt{2}}$。但是，随机噪声并非是正弦的；受带宽限制的噪声的包络其实是呈瑞利分布的，导致平均值会是均方根值的约 1.25 倍。所有这些分析都指出我们应该在测得值上加上约 1.05dB，以便针对检测器特征进行校正。总的来说，为了补偿对数放大器和包络检测仪的偏向性，必须对功率的原始测量值加上 2.5dB。

19.3.1　分辨度带宽与视频带宽的比较

探测仪前后均有滤波器的出现，这让很多频谱分析仪的用户搞糊涂了。IF 滤波器确定

---

⑬ 这个名称来源于雷达发展的早期年代，在那里检测器的输出类似地驱动视频显示器。

一个带宽，在其间噪声功率在被检测到之前被积分。结果是，这个积分值决定了被显示的噪声功率。将分辨度带宽缩小10倍，会使得噪声功率的显示值减小10dB。

另外一方面，视频带宽则决定了叠加在探测器输出上的噪声的总量。由于我们从检测器出来的所希望的信号都是直流成分，那么视频带宽越窄，显示出的电平其噪声就越少。这里要做的折中是，当视频带宽变小时，其建立时间就增加了，而迫使扫描率变慢。因此，分辨度带宽决定了测量值的噪声，而视频带宽则降低了显示出来的噪声的波动程度。

19.3.2 跟踪振荡发生器

有些频谱分析仪允许外部振荡器分享分析仪内部VCO使用的控制电压。这个分立的振荡器，又被称为跟踪振荡发生器，因为它的频率是跟踪着分析仪内部VCO的频率的。跟踪振荡发生器可以被用作被测试设备的信号源，从而允许对DUT频率响应的测量。因此，频谱分析仪在添加了跟踪振荡发生器后就变成为标量网络分析仪(之所以称为标量是由于忽略了相位的考虑，因而不包括全矢量测量)。通过这样一种测量装置，可以快速而方便地对滤波器和放大器的频率响应幅度进行测量。因此跟踪振荡发生器是一个相当有用的配件，大多数的商用频谱分析仪都允许使用一个可选的跟踪振荡发生器。

19.3.3 带有提醒性的附加说明

了解每一个仪器的性能局限总是很重要的。这一点在数字时代更为如此，因为仪器很便利地提供有许多有效位数字的数据，而不管数据内在的合理性。只是因为你付了钱来得到所有这些数位，并不意味着它们都是可以相信的。

一个基本的要求是去观察分析仪的动态范围极限。如果输入信号过大，那么分析仪的有些部分就会超载，造成仪器内部的失真。这种乱真信号的响应可能会被错误地归咎于DUT，使得工程师浪费大量的时间去寻找原本并不存在的问题。幸运的是，频谱分析仪的超载很容易被发现和避免。如果增大衰减倍数(即减小灵敏度)会引起假的频率信号在相对功率上发生变化(或者彻底抵消)，那么频谱分析仪就是问题的根源。作为一个例行步骤，应该通过改变输入灵敏度来验证测量结果确实是归自于DUT，而不是频谱分析仪。

同样地，对允许的输入功率的最大值的了解也是很重要的。超过了这个绝对的上限会造成对前端部件的损坏。这个问题在表征低电平放大器时几乎不会出现。但是，当测试的对象是功率放大器时，就有必要使用合适额定功率的外部衰减器了。另一方面，如果输入信号太小，则DUT输出的信号会被分析仪的背景噪声畸变(甚至完全被淹没)。对分析仪本身的噪声底限几分贝以内的信号的测量一定要多加小心。

另一个要考虑的问题是有用的分辨度带宽为分析仪的内部振荡器的相位噪声所限制。设计得妥善的仪器是“自知的”，即IF滤波器被选择来具有与VCO特征相一致的最小带宽。遗憾的是，并不是所有的仪器都是妥善设计的，所以，通过仔细研究说明书来了解仪器的真正的局限是很重要的。

在从显示屏上读取噪声电平时，重要的是不仅要对IF滤波器，而且对对数放大器以及包络探测仪的噪声带宽都要进行修正。正如前面所叙述的那样，有些仪器包括有游标选项来对这些效应进行自动修正，但并不是所有的仪器都具有。对于后者，必须手动进行修正。

最后，要注意到分析仪LO的谐波失真会造成对单一频率输入的多重响应。这种失真

可以通过仔细的设计被保持在非常小的程度内，但是不能完全消除。对于大输入功率，这种失真会造成明显超过噪声底限的多重响应，从而可能会让人误以为是 DUT 的问题。又一次，对特定仪器特征和局限的仔细了解能帮助用户不为这些假象所迷惑。

19.4　参考文献

下面这些讲义提供了更多的有价值的信息：

"Spectrum Analyzer Measurment Seminar"（频谱分析仪测量讲座）（Hewlett-Packard, February 1988）；

"Spectrum Analysis … Spectrum Analyzer Basics"（频谱分析…频谱分析仪基础，Hewlett-Packard Application Note 150, April 1974）.

Mark Kahrs 的精彩论文"50 Years of RF and Microwave Sampling（射频及微波采样 50 年）"（IEEE Trans. Microwave Theory and Tech., v. 51, no. 6, June 2003, pp. 1787-1805）就采样这个专门课题提供了极为精彩的历史回顾。这篇论文给出了超过 250 篇文章的参考文献，使得这篇回顾文章以非同寻常的透彻深度涵盖了这个课题。

第20章 射频功率放大器

20.1 引言

在本章中，我们考虑将 RF 功率有效与线性地传递到负载的问题。简单的、按比例放大的小信号放大器从根本上说能达到高效率，因此必须考虑其他的方法。与通常的情况一样，这会涉及到各方面的综合考虑，但这次是在频谱纯度(失真)和效率之间进行折中。

在不断寻求提高信道容量的努力下，越来越多的通信系统同时采用幅度和相位的调制。伴随着这一趋势而来的是要求具有更高的线性度(可能同时在幅值域和相位域)。同时，要便携式的趋势又带来了对效率的不断增加的要求。各种各样的功率放大器电路拓扑结构反映出任何单个电路都不能满足所有的要求。

小信号与大信号工作区域对比

回顾一下，在分析包含非线性器件(比如说晶体管)的电路时，做了一个重要的折中。即是说，为了具有用一个线性网络表达，比如说，一个本征上是指数律器件的能力，我们必须接受这个模型仅仅适合于"小"信号这样一个前提。回想一下什么是"小"，以及定量地界定"小"和"大"的边界是有指导意义的。

在接下来的部分，将信号分解为直流部分和信号部分，为了区分哪个是哪个，使用以下记法规范：直流变量用大写字母(及大写字母下标)，小信号部分用小写字母(及小写字母下标)；直流和小信号分量的集合用小写字母变量和大写字母下标的组合表示。

以常见的指数律 v_{BE} 开始：

$$i_C = I_s e^{v_{BE}/V_T} \tag{1}$$

采用传统的做法，在此忽略了 -1 项。

现在，将基极-发射极电压表示成直流和"小信号"成分之和：

$$v_{BE} = V_{BE} + v_{be} \tag{2}$$

然后把上式代入指数律公式，并作级数展开：

$$i_C = I_s e^{v_{BE}/V_T} = I_s\left[1 + \frac{v_{be}}{V_T} + \frac{1}{2}\left(\frac{v_{be}}{V_T}\right)^2 + \cdots\right] \tag{3}$$

在通常的增量分析中，我们仅保留直流和一阶项部分。为了看看高阶项何时忽略比较合适，可以计算出 v_{be} 的一个值(基极-发射极电压增量)，使得其引起的两次项分量不大于一次项的 1/10：

$$\frac{1}{2}\left(\frac{v_{be}}{V_T}\right)^2 \leqslant \left|\frac{v_{be}}{10V_T}\right| \tag{4}$$

解得 v_{be} 为

$$|v_{be}| \leqslant \frac{1}{5}V_T \tag{5}$$

在室温下，v_T 大概为 25mV，因此最大允许的基极-发射极电压偏差量为小得可怜的 ±5mV，对应于集电极电流有相当于静态电流大致 ±20% 的变化。甚至更小的偏差才可能使得两次项分量小于一次项分量的 10%。因此，你可以体会到，"小信号"的确意味着小的信号。

对场效应管的类似推导可以发现小信号是相对于栅过驱动电压，$V_{GS}-V_T$，此处 V_T 为阈值电压，定义的。从绝对的电压值来看，场效应管的线性范围要比双极型晶体管大。然而，这个表面上的优越性是伴随着跨导降低的代价的。

在典型功率放大器中的晶体管会横跨所有的工作区域。因此基极-发射极（或者栅极-源极）电压摆动通常会超过我们刚刚定义的小信号限制。因此，放大器会显示明显的失真，这必须要在分析和设计过程中加以考虑。在窄带放大器中，失真可以通过一个足够高 Q 值的滤波器仅容许通过载波和调制边带信号的方法非常方便地消除。在宽带放大器中，需要采用其他的办法，比如说负反馈和前馈，但通常以降低效率和增加复杂度为代价。

20.2　传统功率放大器拓扑结构

有四种可以被称为"经典"的功率放大器类型，它们的主要差别在于偏置情况不同。它们之所以被称为是"经典"的，是因为在历史上存在在先。这些放大器标称为 A 类、AB 类、B 类和 C 类，并且所有这四种放大器都可以通过分析图 20.1[①] 所示的单一模型来理解。该模型不是唯一的，但是具有代表性。

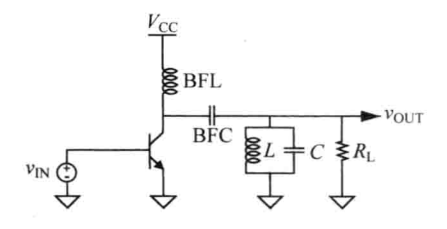


图 20.1　具有经典电路拓扑结构的通用功率放大器模型

在这个通用模型中，电阻 R_L 代表将要把输出功率传递到哪里去的负载电阻。一个"大而胖"的电感 BFL 把 DC 功率送入到晶体管的集电极且假设这个电感很大足以使通过它的电流基本不变。集电极通过电容 BFC 连至一个振荡回路以防止在负载中有任何 DC 功耗。这种特定形式的功率放大器的一个优点是与在传统的小信号放大器中一样，晶体管的输出电容可以被吸收进振荡回路。另一个优点是由振荡回路提供的滤波功能削减了由总是存在的非线性引起的频带外的发射。这种考虑显得特别重要，因为我们不再把自己局限在小信号工作情况下，因而必然会预见到某些失真。为了简化分析，假设振荡回路的 Q 值足够高，因此振荡回路两端的电压即使是由非正弦电流提供，也可以很好地近似为正弦波形式。

20.2.1　A 类放大器

A 类功放的特点是其偏置的选择确保晶体管在整个周期内都导通。A 类功率放大器与小信号放大器之间的主要差别是在功率放大器（PA）中的信号电流是偏置电流的很大一部分，因此可以预见到可能存在的严重失真。A 类放大器是以效率为代价来减弱这个失真，因为即使没有任何信号也总存在由于偏置电流造成的功耗。为了定量地理解为什么效

① 有关这个议题，存在着许多不同的版本，但所有版本的操作特点仍然可以用该模型来理解。

率是差的，假设集电极电流可以非常合理地近似为

$$i_C = I_{DC} + i_{rf}\sin\omega_0 t \tag{6}$$

式中，I_{DC}是偏置电流，i_{rf}是集电极电流中信号分量的幅值，而 ω_0 是信号频率(也是振荡回路的谐振频率)。尽管忽略了失真，但所引起的误差不会严重到使以下的推导不成立。

输出电压就是信号电流与负载电阻的乘积。由于那个“大而胖”的电感 BFL 使基本上恒定的电流在其上通过，KCL(基尔霍夫电流定律)告诉我们，信号电流就是集电极电流中的信号分量，因此，

$$v_o = -i_{rf}R\sin\omega_0 t \tag{7}$$

最后，集电极电压是 DC 集电极电压与信号电压的和。这个“大而胖”的电感 BFL 相当于 DC 短路，所以集电极电压对称地以 V_{CC}为中心摆动[②]。因此集电极电压和电流是相互之间的相位差为 180°的正弦信号，如图 20.2 所示。

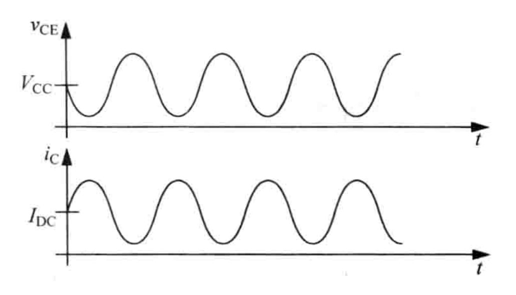


图 20.2 理想 A 类放大器的集电极电压和电流

如果从这些公式中看上去不清楚的话，那么从图中应当能清楚地看出晶体管总是消耗功率的，因为集电极电流和集电极电压的乘积总是正的。为了定量地估计这个功耗，在计算效率时首先计算出传递给电阻 R 的信号功率：

$$P_{rf} = i_{rf}^2 R/2 \tag{8}$$

接下来，计算提供给放大器的 DC 功率。假设使静态集电极电流 I_{DC}正好大到足以保证晶体管从不会发生截止，即

$$I_{DC} = i_{rf} \tag{9}$$

所以 DC 输入功率为

$$P_{DC} = I_{DC}V_{CC} = i_{rf}V_{CC} \tag{10}$$

RF 输出功率与 DC 输入功率的比就是一个效率的度量(通常称为集电极效率)，它可以表示为

$$\eta \overset{\text{def}}{=} \frac{P_{rf}}{P_{DC}} = \frac{i_{rf}^2(R/2)}{i_{rf}V_{CC}} = \frac{i_{rf}R}{2V_{CC}} \tag{11}$$

现在我们注意到，乘积 $i_{rf}R$ 能够具有的最大绝对值是 V_{CC}，否则晶体管将在负值最低峰处饱和。因此，理论上最大的集电极效率就是 50%。如果假设以下这些可能发生的情况：非零的饱和电压，非理想的驱动信号幅度，在谐振糟路和互连线中存在不可避免的损失，那

② 这不是一个打印错误。集电极确实会摆到正电源电压之上。解释这种情况的一种方法是注意到一个理想的电感不可能在它两端出现任何 DC 电压(否则最终会有无穷大的电流流过)。因此，如果集电极电压摆到了电源电压之下，它必然也会摆到电源电压之上。这种思维方式在推导各种类型开关模式功率变换器的特性时特别有用。

么效率常常会得到明显比 50%小的效率——特别是在较低电源电压的情况下尤为如此，这里 $v_{CE,sat}$表示 V_{CC}的较大一部分。因此对于实际的 A 类放大器，集电极效率为 30%～35%是很普遍的[③]。

除了效率之外，另一个重要的考虑是输出晶体管能承受的电学应力。在 A 类放大器中，最大的集电极至发射级的电压为 $2V_{CC}$，而峰值集电极电流值为 $2V_{CC}/R$，因此器件一定要能承受这样大的峰值电压和电流，尽管这两个最大值不会同时发生。

对器件这种承受能力的常用的量化方法是定义另一个品质因子，称为“归一化的功率处理能力”，它简单地就是实际的输出功率与最大的器件电压和电流乘积的比。对这类放大器，这一无量纲的品质因子最大值为

$$P_N \overset{\text{def}}{=} \frac{P_{rf}}{v_{CE,max} i_{C,max}} = \frac{V_{CC}^2/(2R)}{(2V_{CC})(2V_{CC}/R)} = \frac{1}{8} \tag{12}$$

这个品质因子又叫做利用率因子(utilization factor)。

可见，A 类放大器提供的合适的线性度是以低效率和器件相对大的器件承受的电应力为代价的。基于这个原因，A 类放大器在 RF 功率应用中很少见[④]，并且在音频功率应用中也比较少见(在较高功率的情形下尤其如此)[⑤]。

重要的是要再次强调 50%的效率值代表着一个上界。如果集电极摆幅小于前面所假设的最大值，或者在其他地方还有另外的损失，那么效率还会降低。当摆幅接近零时，集电极效率也接近零，因为传递到负载的信号功率变为零而晶体管仍然在消耗 DC 功率。

对这种拓扑结构最后需要讨论的是，尽管放大器因为大信号工作而会发生失真，但由于输出的基波是正比于驱动电流的基波分量，这种放大器还是有线性调制特性。只要驱动电流的基波本身是正比于所希望的调制，在负载两端的也是正比于调制的，这样就得到了线性调制的结果。所以，放大器在描述函数(见第 15 章)这个意义上来说仍然是线性的。

20.2.2 B 类功率放大器

关于如何可以达到比 A 类放大器更高效率的思路实际上隐含在图 20.2 所示的波形图中。这一点应该是清楚的，即如果偏置能设成使漏极电流和漏极电压同时不为零时在一个周期中所占的部分减少，那么晶体管的功耗就会减小。

在 B 类放大器中，偏置设成使输出器件在每个周期的一半时间内关断。当然，一个精确的 50%的导通占空比只是一个数学上的某个点，因此真正的 B 类放大器实际上并不存在。但这个概念有助于我们进行分类。在任何情况下，只要有间断导通，总可以看到严重偏离线性工作状态的情况。但必须注意，输出失真(非线性的标志)以及在输入和输出功率(以基波计算)之间是否存在比例关系(或缺少这一关系)是不同的。一个单端 B 类放大器可能产生非正弦输出，但就这个输入-输出功率的比例关系说它仍然可以被认为是线性工作。当然，我们仍然关心带外的频谱分量，因此必须要求有一个高 Q 的谐振器以得到能接受的近

---

③ 另一个因素是相对失真随着输出功率而降低。在那些低失真是非常重要的应用中，经常用效率换取线性度，这样就导致了十分低的线性度。

④ 也许在低功率水平的应用中或在级联放大器中的前几级是例外。

⑤ 当然，对高端的音频爱好者是例外，对他们而言功耗常常不是一个约束条件。最近，第三代无线通信(3G)对于线性度的要求导致 A 类放大器又开始重新被利用。

似的正弦输出。然而，尽管这个失真，B 类放大器能具有线性调制特性，这又一次是在描述函数这个意义上而言的，因此这类放大器可以被用在要求线性放大的场合下。

尽管这里将要分析的是单个晶体管的 B 类放大器，但应当提及的是大多数实际的 B 类放大器都是两个晶体管的推挽式结构(将在后面进一步讨论这个问题)。

因此对于这个放大器，假设集电极电流在半个周期中是正弦的而在另半个周期中为零：

$$i_C = i_{rf}\sin\omega_0 t, \quad i_C > 0 \tag{13}$$

输出谐振回路滤去了这个电流的谐波，如在 A 类放大器中那样留下的是正弦集电极电压。因此，集电极电流和集电极电压近似地如图 20.3 所示。

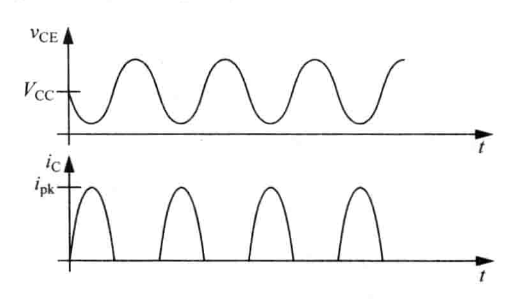


图 20.3　理想 B 类放大器的集电极电压和电流

为了计算输出电压，首先求出集电极电流的基波分量，然后把这个电流乘以负载电阻：

$$i_{fund} = \frac{2}{T}\int_0^{T/2} i_{pk}(\sin\omega_0 t)\ (\sin\omega_0 t)\,dt = \frac{i_{pk}}{2} \tag{14}$$

这个结果是将输出波形削去一半后所预期到的。这样

$$v_o \approx \frac{i_{pk}}{2}R\sin\omega_0 t \tag{15}$$

由于 v_o 最大可能的值是 V_{CC}，由公式(15)中可以很清楚地看出 i_{pk}的最大值为

$$i_{pk,max} = \frac{2V_{CC}}{R} \tag{16}$$

因此峰值集电极电流和最大输出电压与 A 类放大器相同⑥。

我们像前面那样计算集电极效率，首先计算输出功率为

$$P_o = \frac{v_{rf}^2}{2R} \tag{17}$$

式中，v_{rf}是负载电阻两端信号的幅值。幅值的最大值仍然为 V_{CC}。所以最大的输出功率为

$$P_{o,max} = \frac{V_{CC}^2}{2R} \tag{18}$$

计算 DC 输入功率有点复杂，但是还是直截了当的。平均集电极电流为

$$\overline{i_C} = \frac{1}{2\pi}\int_0^{\pi} \frac{2V_{CC}}{R}\sin\Theta d\Theta = \frac{2V_{CC}}{\pi R} \tag{19}$$

所以提供的 DC 功率为

⑥　半个正弦波电流脉冲的假设自然是一个近似。在实际电路中的集电极电流有所不同，主要是过渡到零电流和从零电流起的过渡不是突变的。因此与理想理论所预计的相比，实际器件的功耗偏大一些并且效率偏低一些。

$$P_{DC} = V_{CC}\,\overline{i_C} = \frac{2V_{CC}^2}{\pi R} \tag{20}$$

最后，B 类放大器的最大集电极效率为

$$\eta = \frac{P_{o,max}}{P_{DC}} = \frac{\pi}{4} \approx 0.785 \tag{21}$$

因此，理论的集电极效率明显高于 A 类放大器。然而，再一次，任何实际实现的 B 类功率放大器的真实效率会由于我们已忽略的效应而比这一分析给出的值低一些。然而有一点仍然是正确的，即在所有其他因素保持相同的情况下，B 类放大器能提供明显高于相应的 A 类放大器的效率。

这个放大器归一化的功率处理能力为 1/8，与 A 类放大器相同，这是因为它们的输出功率、最大集电极电压和最大集电极电流都相同[⑦]。

采用 B 类放大器，我们接受了这样一个事实，即用一些失真(但是依然保留了调制的线性性)以换取效率的明显改善。由于这种交换的实现是通过减少一个周期中晶体管导通电流的时间完成的，因此读者很自然会问是否有可能通过进一步减小导通角来达到进一步的改善。对这一设想的探索产生了 C 类放大器。

20.2.3　C 类放大器

在 C 类功率放大器中，偏置被设成使晶体管在小于一半的时间内导通。因此，集电极电流是由周期性的一串脉冲构成的。传统上用正弦波的上面部分来近似这些脉冲以便于进行直接分析[⑧]。具体地说，我们假设集电极电流具有如下的形式：

$$i_C = I_{DC} + i_{rf}\sin\omega_0 t,\quad i_C > 0 \tag{22}$$

式中，偏差值 I_{DC}(其类似于线性放大器中的偏置电流)，实际上对于 C 类放大器来说是负值。当然，总的集电极电流总是为正或为零。也就是说，集电极电流在晶体管导通时是一段正弦波，而在晶体管截止时为零。仍继续假设晶体管在任何时候的工作情况都像是一个电流源(具有高的输出阻抗)[⑨]。

我们也继续假设输出谐振槽路有一个足够高的 Q 值，所以在负载两端的电压基本上保持为正弦波，因此集电极电压和集电极电流如图 20.4 所示。因为我们对是什么将 C 类功率放大器从 B 类功率放大器区分出来有兴趣，让我们的注意力集中在集电极电流脉冲变的相对来讲是窄的时的行为。在这种情况下，通过将脉冲理想化为拥有相同峰值和脉宽的三角波来大为简化这个问题，如图 20.5 所示。

集电极电流的平均值就是三角脉冲面积除以其周期：

$$\overline{i_C} = \frac{1}{T}\int_0^T i_C(t)\,dt \approx \frac{\frac{1}{2}(2\Phi)(i_{PK})}{2\pi} = \frac{\Phi i_{PK}}{2\pi} \tag{23}$$

---

⑦　两个晶体管的推挽 B 类放大器的归一化功率处理能力是这个值的两倍。

⑧　见，比如 Krauss，Bostian 及 Raab，*Solid-State Radio Engineering*(固态无线电工程)，Wiley，New York．1981。他们提供了比在这里表述的更加完整的分析。

⑨　违背这个假设会导致极其复杂的情况。最大效率一般在输出功率接近饱和时获得。在这种情况下，C 类放大器会迫使双极型晶体管在一个周期的一小部分时间内进入饱和区(对 MOS 管则是进入三极管区)，使得精确的分析变得困难。

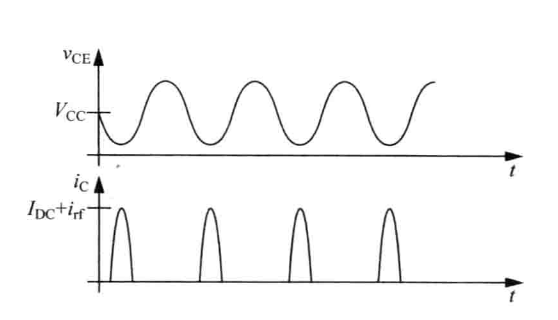


图 20.4　理想 C 类放大器的集电极电压和电流

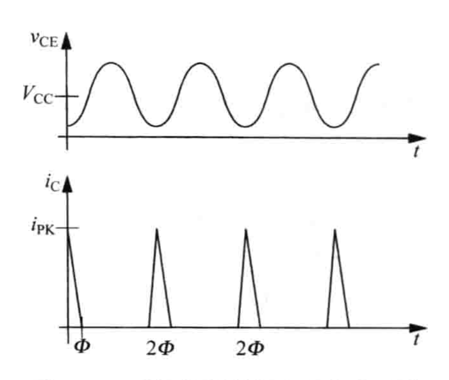


图 20.5　理想化了的集电极电流波形

因此由 V_{CC} 提供的平均功率为

$$P_{DC} = V_{CC}\,\overline{i_C} \approx V_{CC}\,\frac{\Phi i_{PK}}{2\pi} \tag{24}$$

现在，集电极电流的基波分量幅度为

$$i_{fund} = \frac{2}{T}\int_0^T i_C(t)\cos\omega t\,dt \approx 2\,\overline{i_C} = \frac{\Phi i_{PK}}{\pi} \tag{25}$$

在这里我们已经假设这个脉冲足够窄以至于余弦函数在脉冲延续的期间可以很好地用 1 近似。注意到公式(25)的含义，独立于脉冲波形，电流的基波分量总是直流电流的 2 倍——只要集电极电流是一些窄脉冲。这样，在这种情况下，集电极电流波形被理想化为三角形脉冲不会导致严重的错误，但同时能足够简化这个分析过程以利于获得有用的设计内在见解。

一个最重要的内在见解是由公式(25)暗含的非线性。输出电流峰值随着输入电压增加，导通角也是如此。没有理由来期待这两个增加的乘积也是线性的，确实这个乘积不是线性的(事实上，输出幅度是非常近似依赖于输入幅度的平方值)。这样，不同于 A 类和 B 类功率放大器，输出功率非线性地依赖于输入功率。

传输到负载的信号功率为

$$P_{rf} = \frac{i_{fund}^2 R}{2} \approx \frac{\left[i_{PK}\,\dfrac{\Phi}{\pi}\right]^2}{2} \tag{26}$$

$$\eta = \frac{P_{rf}}{P_{DC}} \approx \frac{\dfrac{\left[i_{PK}\,\dfrac{\Phi}{\pi}\right]^2 R}{2}}{V_{CC}\,\dfrac{\Phi i_{PK}}{2\pi}} = \frac{\left[i_{PK}\,\dfrac{\Phi}{\pi}\right]R}{2\pi} = \frac{i_{fund}R}{V_{CC}} \tag{27}$$

现在，最大输出功率是在下式条件下得到：

$$i_{fund,max} = \frac{V_{cc}}{R} \tag{28}$$

因此，如果有足够大的峰值电流和(窄)导通角之积，集电极效率接近于 100%。我们从公式(23)与公式(25)可以看出，如果输出功率恒为常数，调节基极的输入以产生更窄的集电极电流脉冲需要按比例增加集电极峰值电流。这样就可以清楚看出，导通角接近于 0 时，

C 类功率放大器的归一化功率处理能力也接近于 0，这是因为此时峰值集电极电流趋近于无穷大。因为这个原因，实际上导通角并不会接近 0。

尽管之前的处理把导通角作为一个独立变量对待，更普遍的情况是导通角是选择一个很容易实现的输入偏置（比如说零电压）和可用到的输入驱动结合的结果。考虑到这个事实——加上平衡增益、效率和功率处理能力——通常导致导通角在 135°～150°之间[10]。同样，进行上面那些推导的最主要的目的是建立起对设计非常有用的某些一般性的直观感觉——效率可以做得很高，但这是以减少功率处理能力、增益和线性度为代价的。

20.2.4　AB 类放大器

我们已经看到 A 类放大器在 100%的时间里导通，B 类放大器在 50%的时间里导通，而 C 类放大器则在 0 和 50%之间的某段时间内导通。顾名思义，AB 类放大器是在一个周期的 50%和 100%之间的某段时间内导通，这取决于所选择的偏置大小。它的效率和线性度在 A 类和 B 类放大器之间。这种折中常常是令人满意的，这从此类功率放大器的使用非常普遍的事实就可以看出。

20.2.5　D 类放大器

经典的功率放大器都采用有源器件作为受控电流源。另一种方法是器件被用作开关。这个替代方法颇具吸引力，这是开关理想情况下并不消耗任何功率。在任何一个给定的时刻，或者开关两端的电压为零，或者通过它的电流为零。因为开关的 V 和 I 乘积因此总是为零，所以晶体管不消耗任何功率并且其效率必定为 100 %。

利用这一想法的一类放大器是 D 类放大器。初看起来（见图 20.6），它好像就是一个推挽式、变压器耦合的 B 类放大器。与我们通常见到的并联振荡槽路不同，在这类放大器的输出端采用了一个串联的 RLC 网络，这是因为开关模式的放大器与前面介绍过的电流模式的放大器相对偶，所以输出滤波器也互为对偶。

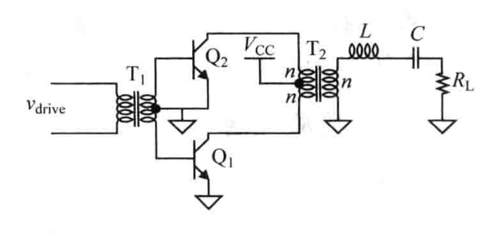


图 20.6　D 类放大器

输入的连接方式保证了在某个给定时间只有一个晶体管被驱动，其中一个晶体管在正半周工作而另一个在负半周工作，就像在推挽式 B 类放大器中那样。这里所不同的是晶体管被深度驱动使它们的工作在开关状态而不是作为线性（或准线性）放大器。

由于开关作用，输出变压器 T_2 的每个初级终端被交替地驱动到地，从而在初级绕组（因而也在次级）两端产生一个方波电压。当一个（晶体管的）集电极为 0V 时，变压器的作用使另一个（晶体管的）集电极电压为 $2V_{CC}$。输出滤波器则只允许这个方波的基波分量流入负载中。

由于在次级电路中只有基波电流流动，所以初级电流也是正弦的，因此每个开关在它导通的半周期中见到的是一个正弦波，所以变压器的电流和电压如图 20.7 和图 20.8 所示。

---

[10] 聪明的读者会注意到，在已给出的推导中所用到的假设这个意义来说，这个导通角的范围可不能算小。然而，总的、定性的结论依然是有效的。

由于晶体管的作用如同开关(至少在原则上是如此)，所以D类放大器的理论效率为100%。

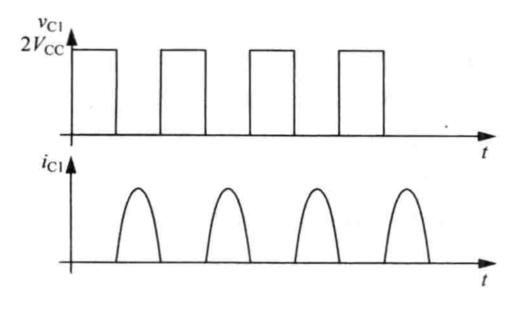


图 20.7 理想D类放大器 Q_1 管的集电极电压和电流

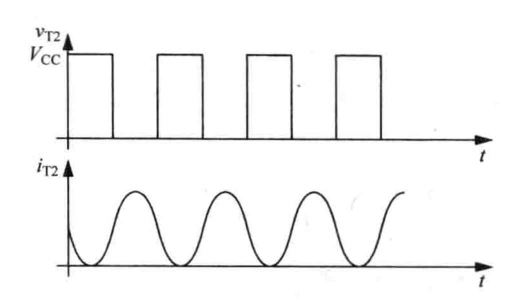


图 20.8 理想D类放大器 T_2 次级的电压和电流

这个放大器归一化的功率处理能力恰为[11]

$$\frac{P_o}{v_{CE,on} i_{C,pk}} = \frac{1}{\pi} \approx 0.32 \tag{29}$$

它比B类推挽放大器要好，并且比A类放大器要好得多。当然，D类放大器通常不能提供线性调制，但它却有可能提供高效率并且不会对器件产生过高的电学应力。

这类(或任何其他开关类型)功率放大器的一个实际问题是并不存在像理想开关这样的器件。在开关过程中非零的饱和电压肯定会产生静态功耗，而有限的开关速度也意味着开关的 *V-I* 积在过渡期间不为零。因此，开关模式的功率放大器只有在频率明显低于 f_T 的情况下才能很好地工作。而且，如果应用双极型晶体管时由于晶体管在饱和区的电荷存储使一个晶体管导通之前另一个晶体管不能完全关断，那么效率的降低就会特别严重，变压器的作用然后就会使全部电源电压加在还没有关断的器件上，从而可能导致 *V-I* 的乘积非常大。

20.2.6 E类放大器

正如我们已经讲到的那样，采用晶体管作为开关可能提供有大幅改善的效率，但由于实际开关(晶体管)不是完全理想的，从而使得实现这种可能性并不总是那么容易。

与此相关的功耗将使效率降低。为了防止大的损耗，开关相对于工作频率来说必须非常快。当载波频率很高时，满足这个要求的困难会变得更大。

如果有一种方法可以修改电路使得在切换瞬间附近的一段不为零的时间间隔内开关两端的电压为零，那么功耗就会降低。E类放大器采用高阶电抗网络来提供足够的自由度以改变开关电压波形，使它在开关导通时的值和斜率均为零，从而降低了开关损耗。可惜的是它对于关断过渡没有任何作用，而关断过渡的边沿常常是更容易引起问题的，至少在双极型设计中是如此。正如我们后面将要讲到的，另一个问题是E类放大器具有很差的归一化功率处理能力(事实上比A类放大器还差)，因此尽管这种类型的放大器可能有很高的效率(理论上在理想开关时为100%)，但它却要求采用更大尺寸的器件把一定数量的功率传送到负载。如果这些限制对设计者来说不是一个问题，那么E类拓扑结构是可以达到极佳的性能的。

---

[11] 也许值得记住，一个方波基波分量的幅值是 $4/\pi$ 乘以方波的幅值。



E 类放大器的另一个特点是很容易被设计。不同于典型的 C 类放大器，它的实现几乎在设计完成后不需要什么调整，就可以很好地工作。

在进行了先前这番讨论之后，现在让我们来看一下图 20.9 所示的 E 类放大器电路。与在前面的例子中一样，那个 RF 扼流圈（RFC，等同于一个 BFL）提供了一条至电源的直流通路而在 RF 时近似为开路。另外，我们注意到电容 C_1 处于很方便的位置，因为任何器件的输出电容都可以被它吸收。

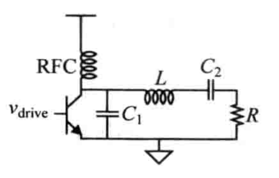


图 20.9　E 类放大器

那些设计公式的推导太复杂了（而且并不提供足够的设计内在了解），在此不多加论述。想要详细了解的读者可以参考如下文献来知道下列公式的推导[12]：

$$L = \frac{QR}{\omega} \tag{30}$$

$$C_1 \approx \frac{1}{\omega(5.447R)} \approx \frac{0.184}{\omega R} \tag{31}$$

$$C_2 \approx C_1\left(\frac{5.447}{Q}\right)\left(1+\frac{1.42}{Q-2.08}\right) \tag{32}$$

尽管我们为输出串联 LC 网络的元件提供了计算公式，但是这种计算只基于简单的、无寄生参数的晶体管模型。实际情况中，就需要找到最优的器件参数以达到最高的效率，同时能够提供所要求的输出功率。

一旦各处都调整好了，集电极电流与电压看上去就如图 20.10 所示。注意，在导通时集电极电压的斜率为零，然而在开关关断时电流却接近最大。因此如果开关不是无限快（正如大多数我们可能遇到的开关的情形），那么在这一过渡期间开关的功耗可以是相当大的。这个功耗足以抵消掉在过渡到“导通”状态时减少的功耗所带来的大部分改善。

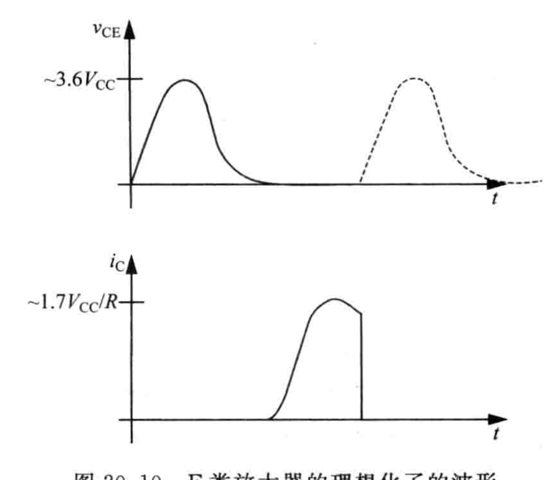


图 20.10　E 类放大器的理想化了的波形

[12] N. O. Sokal and A. D. Sokal，“Class E，a New Class of High-Efficiency Tuned Single-Ended Power Amplifiers”（E 类放大器，一种新的高效率调谐单端功率放大器），*IEEE J. Solid-state Circuits*，v. 10，June 1975，pp. 168-176. 非常感谢 Caltech 的 David Rutledge 教授，是他让我注意 Gerald D. Ewing 的博士论文“High-Efficiency Radio-Frequency Power Amplifiers”（高效率射频功率放大器），Oregon State University，Corvallis Oregon，1964，此文是对这一概念所做的最早的理论说明。

此外，注意每一波形都有相当大的峰值与平均值的比。事实上细致的分析表明，峰值集电极电压近似为 3.6V_{CC}，而峰值集电极电流大致为 1.7V_{CC}/R。

传送到负载上的最大输出功率为

$$P_o = \frac{2}{1+\pi^2/4}\frac{V_{CC}^2}{R} \approx 0.577\frac{V_{CC}^2}{R} \tag{33}$$

因此归一化功率处理能力为

$$\frac{P_o}{v_{CE,pk} \cdot i_{C,pk}} \approx 0.098 \tag{34}$$

正如可以看到的那样，E 类放大器甚至对开关特性提出了比 A 类放大器更高的要求。只要这个特性能够得到满足，E 类放大器可以拥有很优越的性能。

最后，注意到，漏极（或集电极）电流是明显的非正弦波形。因此，几乎可以确定，如果要满足规定标准规定的带外发射，就需要额外的滤波技术。通常，滤波功能和阻抗变换结合在一起（对于正确的输出功率而言）进而/或者部分地吸收到 E 类放大器的电抗性网络元件中去。

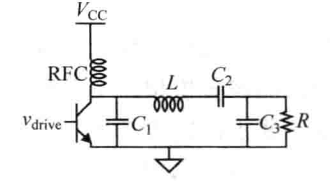


图 20.11　带输出滤波器和阻抗变换的 E 类放大器

图 20.11 中，C_3是组成向下变换 L 匹配网络的一部分。选择低通型的匹配网络可以附带带来滤波特性。那个串联电感 L 有两个作用，完成 L 匹配和作为 E 类放大器的串联 LC 输出网络的一部分。如果需要额外的滤波特性，可以用串联 LC 和输出负载并联以构成陷落（trap）或者在 E 类放大器本身和输出负载之间插入额外的元件来完成。

存在着一个趋势，即人们习惯于将注意力集中在 PA 的输出端，但是我们不应该忽略输入端的驱动条件。特别是在大功率的情况下，功率器件通常都会很大，以至于输入阻抗变得很低（而且和器件的工艺无关）。输入阻抗可以低到欧姆级（电阻加电抗）是不奇怪的，因此，将功率耦合到放大器的输入端在一些场合下颇具挑战性。

最后，对于 E 类放大器而言，双极性晶体管构成的放大器通常没有 FET 的性能好。

这是因为双极性晶体管由于少数载流子而表现出关断延迟效应[13]。E 类放大器的电抗网络只能减缓开启转换时的 V-I 重叠。正如之前提到的，这对于双极性晶体管的更成问题的关断转换没有任何帮助。

20.2.7　F 类放大器

E 类放大器的设计隐含着这样一个概念，即利用电抗终断网络的特性可以将开关电压和电流的波形改变到有利于你目的。或许这个概念最优雅的体现就是 F 类放大器。如图 20.12 所示，这里的输出谐振回路调谐至与载波频率共振，并且假定这个回路具有足够高的 Q 值使它在所希望的带宽外的所有频率上如同短路。

传输线的长度精确地选择为载波波长的 1/4。回想一下 1/4 波长的一段线具有“阻抗倒数”的特性。也就是这样一段线的输入阻抗正比于终端阻抗的倒数：

$$Z_{in} = \frac{Z_0^2}{Z_L} \tag{35}$$

---

[13] 关于这个现象的更仔细讨论，参见与本书一起 CD-ROM 里面包括的补充材料。

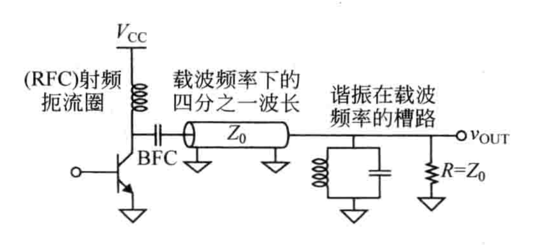


图 20.12　F 类放大器

可以从这个公式中推断出一段长度为半波长的线，其输入阻抗等于负载阻抗，因为两段 1/4 波长部分使我们得到了两次倒数，所以它们互相抵消了。

经过刚才的简短回顾，就可以得到由集电极看到的阻抗的特性。在载波频率，集电极看到的是纯电阻 $R=Z_0$，因为此时振荡回路相当于开路，所以传输线的终端是它的特性阻抗。

在载波的二次谐波处，集电极见到的是短路，这是因为振荡回路在所有远离载波（及其调制边带）的所有频率处相当于短路，而传输线现在就像是半波长的一段线。显然，集电极在载波的所有偶次谐波处看到的是短路，因为在所有偶次谐波处传输线就像是半波长的某个整数倍。相反，集电极在载波的所有奇次谐波处看到的是开路，因为此时振荡回路仍然表现为短路，传输线就像是 1/4 波长的奇数倍，因而提供负载阻抗的倒数。

现在，如果假设晶体管的工作像一个开关，那么这些电抗性的终断保证了集电极电压的所有奇次谐波都将看不到任何负载（除了与晶体管本身输出阻抗有关的负载部分之外），因此在集电极会理想地出现一个方波电压（前面曾讲过一个具有 50％占空比的方波只具有奇次谐波）。

由于在所有高于基波的奇次谐波处传输线都相当于开路状况，所以流入这条线的唯一电流是基波频率的电流。因此当晶体管导通时，集电极电流为正弦电流。而且理所当然，振荡回路保证了输出电压是正弦电压，虽然晶体管只在一半的周期上导通（如在 B 类放大器中那样）。

通过巧妙地安排使方波电压在高于基波的所有频率处看不到任何负载，从而无论在开关接通时刻还是关断时刻开关电流都理想地为零，由此可能达到的高效率可以从图 20.13 所示的波形图中看出。可以看到，总的峰-峰值集电极电压为电源电压的两倍。因此载频处 V_{CE} 基波分量的峰-峰值电压为

$$(4\pi)2V_{CC} \tag{36}$$

注意，由于傅里叶变换的作用，基波的峰-峰值实际上超过了总的 V_{CE} 摆幅。现在由于只剩下基波分量驱动负载，因此提供的输出功率为

$$P_o=\frac{[(4/\pi)V_{CC}]^2}{2R} \tag{37}$$

由于开关没有消耗任何功耗，因此可以得出结论，即 F 类放大器在原理上可以达到 100％的效率。

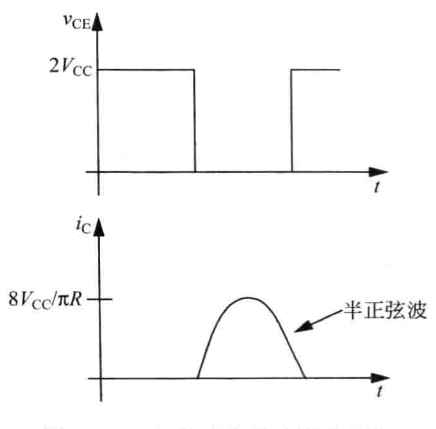


图 20.13　F 类功放集电极电压和电流波形

实际上，可以得到超过 E 类放大器的效率。此外，F 类功率放大器具有好得多的归一化功率处理能力，因为最大电压就是电源电压的两倍，而峰值集电极电流为

$$i_{C,pk}=\frac{2V_{CC}}{R}\cdot\frac{4}{\pi}=\frac{8}{\pi}\cdot\frac{V_{CC}}{R} \tag{38}$$

因此归一化的功率处理能力为

$$\frac{P_o}{v_{CE,pk}\cdot i_{C,pk}}=\frac{\frac{[(4/\pi)V_{CC}]^2}{2R}}{2V_{CC}\left(\frac{8}{\pi}\cdot\frac{V_{CC}}{R}\right)}=\frac{1}{2\pi}\approx 0.16 \tag{39}$$

或者说恰好是 D 类放大器的一半。在某些方面，F 类放大器可以被认为等同于一个单端的 D 类放大器。

应当强调的是 C 类、D 类、E 类或 F 类放大器从本质上讲都是恒包络放大器，也就是说它们通常并不提供一个正比于输入的输出，因此当对它们的全部要求是恒幅值输出（例如这对于 FM 就很合适）时，它们往往工作得最好。但后面会讲到，在要求线性工作的应用中仍然有可能采用这些放大器。此举是很重要的，因为许多现代通信系统所采用的调制都涉及幅值调制（例 QAM）以改善频谱利用率，因此线性工作是必须的。然而目前不得不常常使用 AB 类放大器来满足这个要求，这类放大器相对于恒包络功率放大器来说在效率上会有相应的降低。能够以恒包络的效率实现线性工作的一个通用方法仍然需要进一步探索。我们将在 20.3 节更详细地讨论适用于幅度调制信号的功率放大器的问题。

逆 F 类放大器（F^{-1}）

F 类放大器的对偶本身也是一个在效率上具有相同的理论极限的功率放大器[14]。

F 类放大器的终端在高于载波基波的奇次谐波处表现为开路，偶次谐波表现为短路，而逆 F 类（其简化表示为 F^{-1}）放大器的终端在偶次谐波处表现为开路而在奇次谐波处表现为短路，见图 20.14。

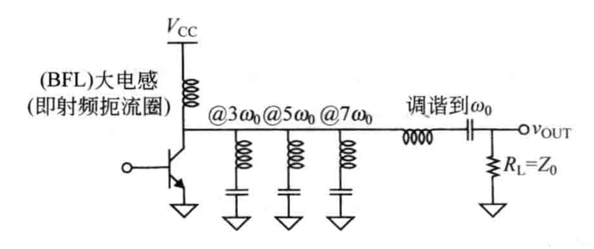


图 20.14　反 F 类放大器（所示的为由三个集总元件谐振支路组成的例子）

同样，如果比较有利或实际可行，也可以用一条传输线来替代集总谐振器。此时，用在基波频率下长度为 λ/2 的一段传输线替代并联在一起的那几个串联谐振器以插入在集电极和输出的串联 LC 槽路之间。

当有无穷多个串联谐振器时，集电极端电压波形表现为理想的（半）正弦波而电流波形则为方波，如图 20.15 所示。我们又一次看到，与 E 类和标准的 F 类放大器一样，当开关切

⑭　见，比如，S. Kee *et al.*，"The Class E/F Family of ZVS Switching Amplifiers"（E/F 类 ZVS 开关放大器），*IEEE Trans. Microwave Theory and Tech.*，v. 51，May 2003.

换时不存在 V-I 之间的重叠，这是这种结构在理论上具有高效率的原因。

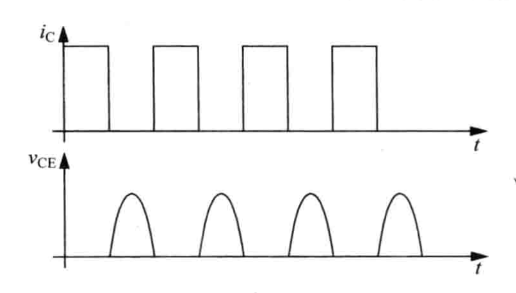


图 20.15　理想反 F 类放大器的集电极电压和电流波形

替代的 F 类拓扑结构

图 20.12 所示的拓扑结构是十分巧妙的，但在许多应用中所要求的传输线可能会不切实际地长。而且在高于基波的奇次谐波处有无穷大(或接近无穷大)阻抗带来的好处常常在实际中会被晶体管本身的输出电容所消除。因此采用集总元件近似常常可以与传愉线工作得几乎一样好。

为了建立这样的集总近似，可以用串联起来的一些并联谐振滤波器来代替传输线。每个谐振器调谐到载波频率不同的奇次谐波上。常常只需将一个振荡回路调谐到 $3\omega_0$ 就已经足够了。在超过图 20.16 所示的两个振荡回路时很少会发现效率有明显的改善。例如，采用一个调谐至三次谐波的槽路可使集电极效率最高提升至 88%左右，而 B 类放大器的效率最高为 78%左右。增加调谐至五次或七次谐波的槽路可使 F 类放大器的效率分别提高至 92%和 94%[15]。但是考虑到实际的槽路元件不是无损的，那么回报递减的规律将很快使采用附加的谐振器变得毫无用处。

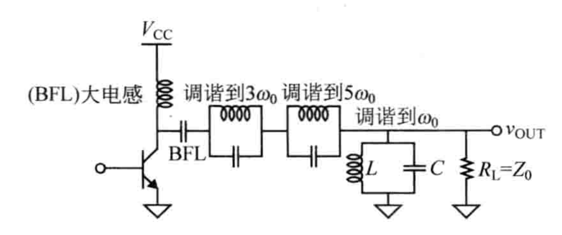


图 20.16　另一种形式的 F 类放大器拓扑结构

20.3　功率放大器的调制

20.3.1　A 类、AB 类、B 类、C 类、E 类及 F 类

调制 A 类或 B 类放大器在原则上是很容易的，因为输出电压直接与集电极电流中信号分量的幅值 i_{rf} 成正比。因此如果 i_{rf} 本身正比于输入驱动，那么就可实现线性调制。采用短

---

[15] F. H. Raab, "Class-F Power Amplifier with Maximally Flat Waveforms"(具有最大化平坦波形的 F 类放大器), *IEEE Trans. Microwave Theory and Tech.*, v. 45, no. 11, November 1997, pp. 2007-2012.

沟 MOS 器件与其他场效应晶体管(FET)可以很容易地实现对这一比例关系的很好近似，因为当有足够大的栅电压时，短沟 MOS 器件具有恒跨导。双极型器件由于有串联基极电阻也能提供合适的线性度，这个串联基极电阻或者由外部提供，或者就是器件本身的电阻。然而，随着无线通信系统的发展，人们对线性度的要求已变得越来越高。一种粗放的(但几乎被普遍采用的)线性化"方法"是功率回退(power backoff)，也就是说，比如，我们从一个能产生 10W 功率的放大器中只要求 1W 的功率[16]。

功率回退的基本原理很容易理解，只要考虑一下在第 13 章中为定义 IP3 而采用同一种弱非线性放大器模型就可以了。由于三阶 IM 项在输入功率每下降 1dB 时下降 3dB，所以输入功率每减小 ldB，基波和三次谐波分量之间的比就会改善 2dB(同样，输入功率每减小 1dB 时，相应的五阶或七阶 IM 项的比在理论上将分别改善 4dB 和 6dB)。如果这种趋势成立，那么就存在某一水平的输入功率，当低于这个水平时输出 IM3(及其他)失真分量相对于载波的功率就会低到可以被接受的程度。由于 A 类放大器的效率和输出功率都将随功率回退数量的增加而减小，因此应当采用相应于达到失真指标时的最小(功率回退)值。典型的功率回退值相对于 1dB 压缩点来说曾普遍低于 6～8dB(并且甚至常常低至 1～3dB)。近年来常常会发现功率回退值必须高至 10～20dB 才能满足某些系统对线性度的严格要求。由于采用分贝表示数值时很容易失去对实际情形的跟踪，因此让我们重新考察一下最后一个数值范围(10～20dB)的含义：这个范围是指，在你费尽心机设计了一个 10W 的 RF 放大器之后，你也许会发现，只有当它的输出功率保持在几百毫瓦(mW)以下时才能满足频谱纯度的指标要求。

与 A 类放大器相比，由于 AB 类放大器的结构本身就有较大的非线性度，所以它的输出 IM 分量表现出与输出功率之间具有较弱的相关性(例如输入功率每减小 1dB，IM3 功率则下降 2dB)。更糟的是，任何数量的功率回退都不能达到可接受的失真的情况并不少见。最后，与 A 类放大器一样，功率回退常常会使效率降低到不可接受的程度(例如在某些情形中低至 5%～10%)。我们很快就会讨论线性化的其他方法，可以用这些方法来缓解某些这种折中关系。

C 类放大器提出了更为严峻的挑战，这可以通过分析前面推导出的输出电流公式来理解：

$$i_{\text{fund}} = \frac{\Phi i_{\text{pk}}}{\pi} \tag{40}$$

尽管看上去会以为有一种线性关系存在，实际上流过电阻负载的电流基波分量与 i_{pk} 之间一般不呈线性，这是因为导电角 Φ 也与 i_{pk} 有关[17]。因此，C 类放大器通常并不提供线性调制能力，所以一般不适合用于幅值调制，至少当被调制的载波驱动基极电路时是如此。

为了从非线性放大器(例如 C 类、D 类、E 类或 F 类)中获得线性幅值调制，把接电源端(漏极电路)考虑成一个输入端比较好。这个通用的想法很简单：改变电源电压就可以改变输出功率。这里的控制实际上可以比在通常的输入端(例如栅)更为线性。首先按这一想法去做

[16] 假若别人采用一种低级、初步的技术可以被称为"凑付"，你自己用这样的技术就可以被称为是一种方法了。

[17] 亦即，除了 A 类或 B 类的工作情况是特例。对于其他情形，在驱动和响应之间并不存在着比例关系，所以线性调制并不是一个固有的特性。

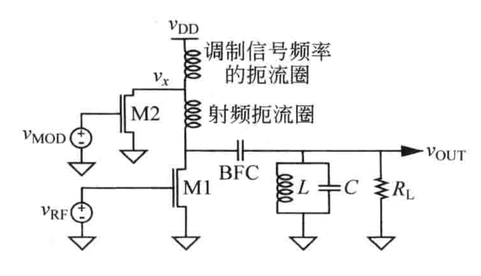


图 20.17 具有 C 类 RF 放大级的 MOS Heising 调制器(简化电路)

的显然就是 AT & T 公司的 Raymond Heising，他于 1919 年左右在(当时自然是)真空管放大器中实现了这一点[18]。Heising 调制器(由于它采用扼流圈，所以也称为恒电流调制器)最简单的 CMOS 实现显示在图 20.17 中。

调制放大器 M2 的负载是一个扼流圈("调制扼流圈")，它选择得足够大，使在最低调制频率时具有很高的电抗。电压 V_x 是 V_{DD} 与在这个调制扼流圈两端产生的调制电压之和。这个电压的和又通过 RF 扼流圈传送到 M1(偏置成 C 类放大器)作为它的等效电源电压，这一点与我们所有传统 PA 结构通常的模型一样。由于两个晶体管共享一个公共的 DC 电源，并且由于电压 V_x 只是接近于地(但不会是地)电压，所以 M1 的输出永远也不会达到零。因此，基本的 Heising 调制器本身是不可能达到 100％的调制深度的(这个特点在某些情形中是一个优点，因为过调制以及随之而带来的严重失真从电路本性上来讲将不可能出现)。在调制放大器产品的实例中，60％～80％的最大调制百分比并不少见。对于要求全调制范围的应用，一个简单的解决办法是将一个有电容旁路的电阻与 RF 扼流圈相串联。在该电阻上的 DC 电压降将使 M1 比 M2 工作在更低的电源电压下。其缺点是调节深度的改善是以效率的降低为代价的，这是因为在这个附加的电阻上存在功耗的缘故。

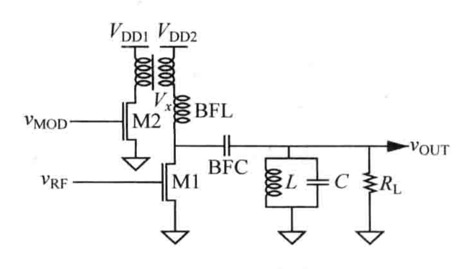


图 20.18 另一种漏极调制例子

此外还有其他形式的漏极调制，它们不需要以牺牲效率来改善调制深度。一个使用广泛的例子显示在图 20.18 中。我们可以利用电源电压、变压器匝数比以及调制幅度之间的多种组合来迫使最大调制时 M1 的漏极电压为零(而不只是接近于零)。例如，假设选择 1∶1 的变压器。由于希望次级电压能达到零，那么初级摆幅的幅值就必须达到 V_{DD2}。因此这种要求就迫使我们选择比 V_{DD2} 大一点的 V_{DD1}(以包括 M2 两端非零的电压降)以及一个合适的栅极驱动电压，以产生所希望的调制摆幅。

更普遍的做法是采用单一一个漏极电源电压，这时要求变压器的匝数比不同于 1∶1。通过选择合适的升压比可以达到 100 ％的调制。

这里需要注意的是调制器本身也是一个功率放大器，因此这些高电平调制器也与它们所驱动的 RF 功放级一样，具有本质上相同的需要在效率和线性度之间做出折中选择的问题。如果不注意，调制器的功耗就可能超过 RF 主功率放大器的功耗。另一个普遍用来解决该问题的电路是用一个推挽 B 类放大级来代替 M2 以提高效率。在这种情形中，B 类放大级的输出漏极连至变压器的初级，后者的中心抽头连至 DC 电源。效率更高的另一种方

---

[18] 见 E. B. Craft 和 E. H. Colpitts，"Radio Technology"(无线电技术)，*AIEE Trans.*，v. 38，1919，p. 328。还可参见 R. A. Heising，"Modulation in Radio Technology"(无线电技术中的调制)，*Proc. IRE*，v. 9，August 1921，pp. 305-322，以及 *Radio Review*(无线电评论)，February 1922，p. 110.

法是通过一级工作在开关模式(例如D类)的放大器产生电压V_x，其要求是必须能够充分滤去开关噪声以满足严格的频谱纯度要求，但它带来的高效率常常证明这种工程上努力是合算的。有时在开关调制器中会采用$\Delta-\Sigma$调制来改变噪声谱的形状以降低对滤波的要求。

这几个例子表明有许多方式可以影响漏极(高电平)调制⑲。然而，虽然漏极调制允许用非线性放大器实现名义上的线性调制，但线性度仍可能不足以满足对频谱纯度的实际要求。这一缺点促使我们去考虑各种增强措施及其他解决办法。

20.3.2 线性化技术

包络反馈

也许把功率回退和漏极调制称为线性化技术有些过于大方了。功率回退以效率换取线性度，然而效率是太宝贵了因而不能随意浪费。漏极调制虽然对某些结构来说优于栅极调制，但它最终仍然要依赖于开环特性，因此设计者并不能直接控制失真。在本小节中，我们将讨论许多以最小的在效率上的代价来提高RF功率放大器线性度的方法。

在面临提高放大器线性度的一般问题时，很自然就会想到负反馈。然而围绕着RF功率放大器加上一个通常的反馈环是非常冒险的。如果采用电阻反馈，那么在大功率放大器中反馈网络的功耗实际上可能非常大，以至于足以引起散热问题，更不用说效率总是随功耗增大而降低了。电抗反馈确实不存在这个问题，但我们却必须格外小心以避免这一电抗可能引起的寄生谐振。此外还有一个环路传输幅值是否充分大的问题，这是所有放大器都存在的问题。非线性度减少的倍数等于环路传输幅值(实际上应当是回路差值，但当环路传输幅值很大时这两个值近似相等)，但其代价是闭环增益也降低了同样的倍数。闭环增益降低10倍将使三阶IM失真(归一至基波)减小10倍，因此需要提供额外的增益才能使线性度有较大的改善。在射频工作时，开环情况下能获得的增益已经来之不易了，因此如果闭环增益不降低到得失并存的地步，则几乎不可能使线性度得到显著的改善。

如果不想付出这么大的代价而是试图采用串联许多放大级的方法，则会出现传统的稳定性问题。随着我们寻求更大的带宽，这一问题将越来越严重，因为寄生极点落在频带内使稳定裕量减少的可能性将加大。

机敏的读者会注意到，实际上只需在调制信号的带宽上进行线性化，并且这个带宽并不需要处于载波的中心频率附近。作为这一观察的具体应用，假设反馈回一个相应于输出信号包络的信号(例如利用一个解调器，它在非关键应用中可以采用简单的二极管包络检测器)，然后用这个解调信号形成闭环⑳。

在基带频率下将环路闭合潜在来讲是有不少优点的，因为这样可以很容易地在所关注的带宽上获得所需要的额外环路增益。但要满足所有这些相关要求未必很容易，特别是如果希望在较大的带宽上线性度有较大改善时。举一个简单的例子就足以清楚地说明这些相关的问题。

假设希望在1MHz的带宽上使失真减少40dB，那么必须在1MHz时有40dB的额外增

---

⑲ 尽管许多文献并不区分漏极调制和Heising调制，但这里要说明后者是前者的子类。

⑳ 例如，见P. E. Terman和R. R. Buss，"Some Notes on Linear and Grid-Modulated Radio Frequency Amplifiers"(有关线性与栅极调制无线电频率放大器的一些见解)，*Proc*，*IRE*，V. 29. 1941. pp. 104-170.

益。如果反馈环可以很好地被建模成单极点系统,那么相应的环路交零频率将是 100MHz,这意味着要求稳定的闭环带宽也是 100MHz。想确保该环路在这一带宽上的特性确为一个单极点系统并非不可能,但的确不是一件容易的事[21]。从这些数字可以很容易看出,如果想在一个更宽的带宽上获得更大的线性度改善,那么困难就会迅速加大。

即便只要求幅值线性度有比较适中的改善,对相位特性的约束(相对于基带)也仍会引起设计困难。我们注意到一个单极点系统在 -3dB 频率处的相位滞后是 45°,如果对这个通带上允许的相移有严格的规定(例如,为了限制群延时的差别),那么对一个单极点系统的唯一解决办法是提高带宽。如果允许的相位误差为 5.7°,那么带宽就应当选择在基带带宽的十倍频以上;如果允许的误差范围缩小到 0.57°,那么所要求的带宽就会再提高一个数量级,即增加到基带带宽的 100 倍[22]。

这些计算都乐观地假定唯一的误差来源来自前馈路径而反馈在各个方面都认为是理想的。这个要求相当于需要一个在很宽的动态范围上有极好线性度的解调器,因为一个负反馈系统只对前馈路径上的非理想性降低灵敏程度。由于整个系统的性能受限于反馈的质量,因此在解调器中的任何非线性和相移都会限制环路的效率。

这些困难足以令人气馁,由此出现了一系列其他技术用来代替或补充传统的负反馈技术。其中有些技术是纯开环技术,因此它们不受稳定性考虑的限制。而且,我们将介绍的那些线性化技术,它既可以单独使用也可以与其他技术联合使用,这取决于具体的设计目标。

前馈

一种开环线性化技术——前馈,是由 Harold Black 在他发明负反馈放大器之前提出的,见图 20.19。

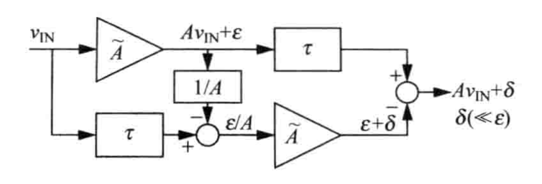


图 20.19 前馈放大器

注意到前馈能够显著改善线性度的带宽部分取决于各个放大器的群延时能够通过可实现的延时元件精确跟踪的带宽[23]。这种跟踪必须在长时间的使用过程中,并且在存在温度和电源电压变化的情况下仍然保持精确。在许多实际产品(如某些 GSM 基站功率放大器)中,延时元件基本上是用适当长度的低损同轴电缆实现的。与大多数依靠匹配的技术一样,可以指望大约有 30dB 的改善(如果十分小心的话也许可以达到 40dB 以上)。在某些情形中还可以实现自动微加工技术,其中一些自动微加工技术依赖于对放大器发送的引导信号。具有这种特性的自动校正技术可以使前馈提供高度一致的优质线性度。如果还必须进一步改善线性度的话,那么可以将这种技术与其他技术联合起来使用。

---

[21] 放宽对单极点的限制也许有助于降低对额外带宽的要求,但却有产生条件稳定的反馈系统的危险。

[22] 我们也可以采用合适的相位补偿全通滤波器(all-pass filter)来降低对额外带宽的要求。

[23] 由于实际的放大器一般不会呈现出恒定的群延时,因此设计一个补偿延时元件来跟踪这种实际行为是十分困难的。

尽管利用前馈可以达到比较宽的带宽，但由于两个完全相同的放大器都消耗功率从而导致了低效率，这无疑是一个缺点。虽然因有两个单独的放大路径而提供部分冗余常常是一个有吸引力和可作为补偿的优点，但由于效率太低（一般在10%以下），因此在大多数应用中往往不采用前馈的RF功率放大器。

预畸变（pre-distortion）和后畸变（post-distortion）

另一种开环线性化的方法是利用以下事实：即把一个非线性元件与它在数学（表示）上相反的元件串联在一起就可以使总的传输特性成为线性。这个补偿元件可以放在非线性放大器的前面或后面，逻辑上可以很合理地分别将它们称为预畸变器（predistorter）或后畸变器（postdistorter）。预畸变至今是这两种变换中采用得更普遍的一种（因为在PA本身的输入端处功率水平较低），并且它可以应用在基带或RF中。基带预畸变极为普遍，这不仅因为基带频率比较低，而且还因为实际上可以选用模拟技术或数字技术来实现（其中数字技术因具有特有的灵活性而变得日益普遍）。基带预畸变的另一个优势是可以纠正在上变频至RF时引起的非线性，见图20.20。

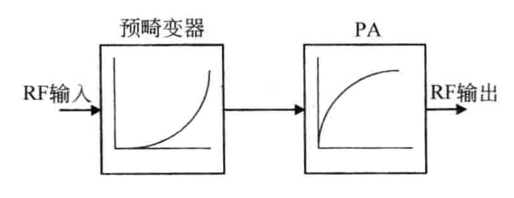


图 20.20　RF 预畸变示意图

由于放大器中主要的非线性与增益压缩有关，因此预畸变器可成功应用的范围只是当输入增加时它能够通过提高增益来精确地抵消增益的压缩时。但是必须记住，一个预畸变器不可能增加放大器的饱和输出功率，因此几乎或根本不应当指望在1dB的压缩点处有任何改善。由于对于“正常”的非线性来说IP3至少在某种程度上与压缩点是相关联的，因此我们不会对为什么预畸变很少能成功地使IM3分量的减少比十来个dB多很多的现象感到奇怪。如果需要大幅度的减少，那么单靠预畸变多半不能成功[24]。

相位误差（包括那些可能由AM至PM转换引起的）的校正可以通过在输入端串联一个移项器来实现。大部分放大器对于小幅值输入往往表现出较大的相位滞后，因此对移项器的控制必须安排以来补偿这个相移。这里的约束通常不如幅值校正那么严格，但设计一个模拟控制电路却非常复杂，因此数字控制已变得十分普遍。为此，一旦要花功夫来构建一个数字控制器时，最好让这个控制器同时控制增益和相位校正器。此举离实现一个同时使幅值和相位路径形成回路的真正反馈环只有一步之遥，一旦闭环形成了就实现了极坐标反馈（polar fcedback），我们很快会对这一课题给出更多的说明。

即使要达到前面提到的十几个dB这一不算十分大的改进——而且不管预畸变是用纯模拟电路还是数字控制单元来实现——我们也必须解决精确产生所希望的反传输特性的问题，并且还要保证这种反特性在面临通常的工艺、电压和温度变化的情况下随时（以及随可

[24] 偶尔也会在一些文献中看到有大幅度的减少，但在仔细考察之后会发现，许多这样的报道都是指那些同时采用功率回退（或其他某种方法）和预畸变的系统，虽然这些系统未必明确承认采用了功率回退。

能变化的负载)保持正确[25]。

因为固定的预畸变器不能恰当地容纳这样的漂移而不发生变化,我们很自然会想到用自适应的预畸变器替代它。这样的预畸变器实时测量,比如电压和温度,来周期性地计算和更新反函数。因此它的成功实现需要有一个系统模型以及测量相应输入变量的传感器。遗憾的是,系统的建模是一件非常困难的事,特别是如果某些重要变量(如输出负载,它在便携式应用中的变化范围可能很大)不能方便地测量(如果能被测量)的话更是如此。使这个困难更为麻烦的是,由于有能量存储,因此非线性可能存在迟滞效应。在这些情况下,当前的输出值不仅与当前的输入而且也与过去的输入值有关。这些局限性并不意味着预畸变没有任何意义(事实上正好相反,因为许多广播电视发射机都依靠这一技术),但它们确实解释了为什么预畸变要持续地使线性度有较大的改善是非常困难的。与其他技术一样,预畸变可以与其他方法结合起来使用以达到总的线性度目标。

包络的消除与恢复

最先由 Leonard Kahn 提出的用来改进单边带(SSB)发射系统的包络消除和恢复(EER)技术就其本身来说并不是一种线性化技术,而是通过漏极调制从非线性(恒包络)放大器中实现线性放大的一个系统[26]。在 EER 技术中,需要被线性放大的调制后的 RF 信号被分成两路,见图 20.21。一路传送到限幅放大器(limiting amplifier)(其本质上是一个比较器)以产生一个恒包络的 RF 信号,这个信号接下来被一个恒包络(例如 C 类)放大器进行高效率的放大。另一路传送到一个包络检测器(解调器),所提取的调制信号再送到采用漏极调制的恒包络放大器。由于 EER 本身并不是一种线性化的方法(最好把它称为效率提升技术),因此,为达到可接受的频谱纯度也许还需要用真正的线性化技术进行补充[27]。

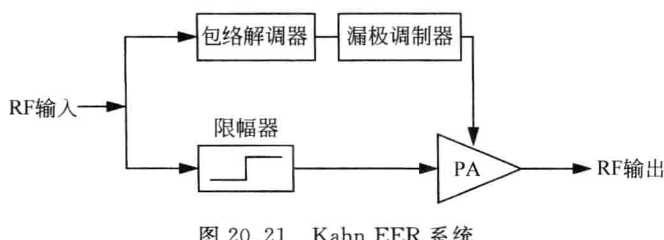


图 20.21　Kahn EER 系统

现在的情况是,建立任何一种理想的元件(特别是在 RF 的情况下)都是很困难的,因此有必要分析一下我们实际上对限幅器究竟有什么要求。在 Kahn 的 EER 系统中,限幅器的作用只是向 PA 级提供合适的驱动以保证高效率的工作。但正因为这样,也许实际上比较有利的是使 PA 的输入随 RF 输入的包络(至少大致如此)变化而不是保持不变,以便在包

---

[25] 预畸变作为一个提供总体线性操作的方法对读者应该是熟悉的。比如说,一个电流镜实际上就是依赖于一对逆向、非线性的变换(首先从电流到电压,然后再变回到电流)以在电流域中提供真正的线性行为。一个更为复杂的例子是真正的吉尔伯特(Gilbert)增益单元,它的工作依赖于同样的一个非线性转换器(I 到 V,然后再变回来)的级联。

[26] L. R. Kahn. "Single Sideband Transmissions by Envelope Elimination and Restoration"(通过包络消除与恢复的单边带发射),*Proc. IRE*, v. 40, 1952, pp. 803-806。顺便说一下,确实应该是 Kahn 而不是 Khan,后者是 Kirk 船长的敌手。

[27] D. Su 和 W, McFarland. "An IC for Linearizing RF Power Amplifiers Using Envelope Elimination and Restoration"(采用包络消除与恢复的线性化射频功率放大器的集成电路), *IEEE J. Solid-State Circuits*, v. 33, December 1998. pp. 2252-2258.

络较小时不需要太大(因而造成浪费)的PA驱动。因而，设计一个实际的限幅器可以大为简化，因为这个问题实质上归结成了设计一个放大器而不是一个严格意义上的限幅器——同时又不需要过多地考虑幅度的线性度[28]。根据具体的PA特性，这一限幅器也许在输入幅值较低时必须提供较高的增益以确保PA级总能被深驱动，从而保证高的效率和较低的噪声底限[29]。同样，也许有必要插入一个补偿延时(一般在RF路径上)以确保漏极调制在时间上能与PA驱动保持一致。如果不能保持一致，就有可能在低功率输入时影响EER正常工作的能力，使输出功率的可用动态范围减小。当然，在实际中达到大约比30dB大得多的动态范围仍然是相当困难的。

Chireix错位相位(RCA Ampliphase)与用非线性元件实现线性放大(LINC)

通过将几个非线性放大器的输出进行组合以获得线性调制的技术被通称为LINC(linear amplification with nonlinear components，用非线性元件实现线性放大)[30]。在文献中首次提出LINC概念的是Henri Chireix(其发音有点像she wrecks)，他在1935年左右提出了相位错位调制(outphasing modulation)。相位错位通过把两个具有不同相位的恒幅值信号向量相加来产生幅值调制[31]。恒幅值特性使我们可以采用高效率的恒包络RF放大器，而向量相加则不再需要漏极调制从而消除了与之相随的功耗。

错位相位调制在发明之后的20年间断断续续地获得了一些商业成功，但它的迅速普及是在RCA著名的AM广播发射机系列采用这一技术时，即起始于1956年的50 kW BTA-50G发射器[32]。Ampliphase这个在RCA市场广告中使用的名字也许就是此后15年间AM广播无线电发射机的主要技术。

为了实现错位相位方法，首先使基带信号从单端变为差分信号(如果还不是以差分形式存在的话)，然后利用这两个输出对一对等时(即同频率)的但非同步的RF载波进行相位调制；接下来由高效率的恒包络放大器放大这两个经过相位调制的RF信号；最后用一个简单的无源网络把这两个放大后的相位调制信号相加。于是经过幅值调制的RF信号就出现在这一合成器上，如图20.22所示。

在典型的实现中，选择这两个放大器输出之间的静态相移ϕ_0等于135°。这两条信号路径设计成每一个可以分别产生45°和−45°的最大相位偏移，因此合成器的两个输入之间的总相位差在90°和180°之间变化。当相位差为90°时，这两个信号相加产生最大的输出；当相位差为180°时，这两个信号相互抵消产生的输出为零。这两种极端情形对应于经过调制后信号的两个峰值，其调制深度即为100%。

---

[28] 我们仍然必须注意仔细设计以免在不允许存在AM至PM转换的通信系统中出现这种转换。这个要求适用于整个放大器，因此它与我们是否寻求实现一个传统的滤波器没有关系。

[29] F. Raab. "Drive Modulation in Kahn-Technique Transmitters"(采用Kahn技术的发射器中的驱动调制)，*IEEE MTT-S* Digest v. 2 June 1999. pp. 811-814.

[30] D. C. Cox. "Linear Amplification with Nonlinear Components"(采用非线性元件的线性放大)，*IEEE Trans Communications*. December 1974，pp. 1942-1945.

[31] H. Chireix. "High-Power Outphasing Modulation，"(高功率电平的错位相位调制)，*Proc. IRE*. v. 23，1935. pp. 1370-1392.

[32] D. R. Musson，"Ampliphase... for Economical Super-Power AM Transmitters"(Ampliphase：经济的超高功率AM发射机)，*Broadcast News*. v. 119，February 1964，pp. 24-29；还可参见*Broadcast News*. v. 111，1961. pp. 36-39。错位相位提升的效率足以建立输出功率至少为250 kW的实用的发射机。

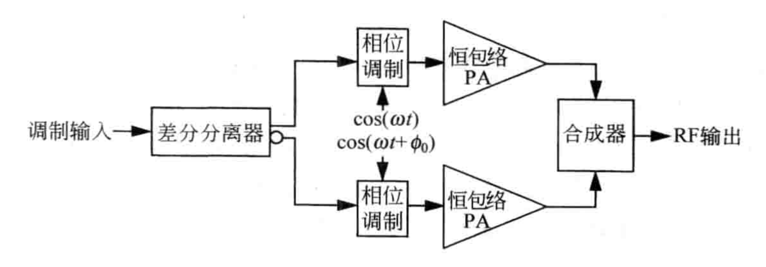


图 20.22　失配相位调制器的方框图

在设计错位相位系统时主要的工作是：(a)获得线性的相位调制；(b)实现一个低损耗合成器，它能避免一个放大器被另一个放大器牵拉(pulling)而降低来之不易的效率、线性度及稳定性。在 Ampliphase 系统中，相位调制器所利用的相位变化是通过改变一个槽路的 Q 值得到的，这个槽路的中心频率与载波频率有些偏差，见图 20.23。

晶体管 M_1 的输出电阻作用为一个可变电阻，它的数值随调制而变化。当调制电压上升时，M_1 的输出电阻下降，因而提高了输出槽路的 Q 值。晶体管 M_2 就是一个跨导器，它把 RF 电压转变成电流。因此，输出电压相对于 M_1 漏极电流的相位角与槽路阻抗的相位角相同(可能会差一个正负符号)。这个相位角又与槽路的 Q 值有关，因而也与调制有关。注意，相移与调制电压之间的线性关系很难用这个电路来保证，因此需要采用预畸变来获得名义上线性的调制。

另一个设计困难来自输出功率的合成网络。在 Ampliphase 发射机中合成器基本上是一对 CLC 的 π 网络(总是需要它来进行阻抗变换)，它们的输出连在一起(每个 π 网络的作用是一个 1/4 波长导线的集总近似)。该情形显示在图 20.24 中。

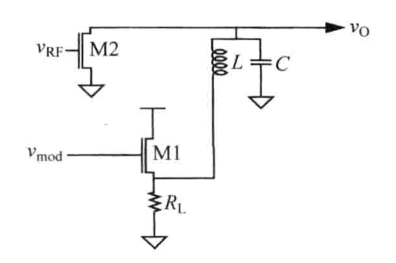


图 20.23　Ampliphase 相位调制器(简化了的 CMOS 版本，具体的偏置电路未画出)

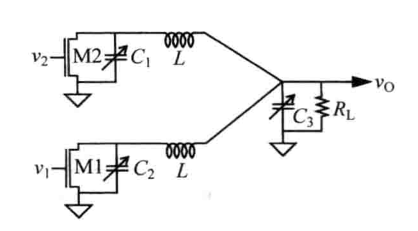


图 20.24　Ampliphase 系统中的输出合成器(简化了 CMOS 版本，偏置电路被忽略)

这个合成器看似简单，但实际并非如此[33]。虽已证明在很成功的产品设计中错位相位是可行的，但它也显现出一个根本问题。从每个晶体管漏极看出去的有效阻抗不仅取决于连至最后输出的负载，而且也取决于另一个晶体管漏极处信号的相对相位。为了可以容忍大范围变化的漏极负载阻抗，电容 C_1 和 C_2 必须随驱动的瞬时相位角变化(在相反方向上)。不言自明，对线性可控电容的这一要求只会加大设计难度。事实上，要求有一个实用的、低功耗和线性的能提供高度隔离的合成器仍然没有实现。主要就是由于这个原因，LINC 今

[33] 如果我们用一段合适的传输线来代替 π 网络，那么它看上去将更简单，虽然仍然需要可变的补偿电容。

天仍然没有成为主流，虽然它的结构看上去很有吸引力。

获得线性相位调制及设置正确的静态相位——更不用说使它们在时间、温度和电源电压变化的情况下仍能保持正确工作——带来的困难，使广播工程师们有时既钟爱又嘲弄地称这些发射器为“Amplifuzz”（fuzz 是麻烦的意思，译者注）。到 20 世纪 70 年代中期，Ampliphase 的产品系列就基本上逐渐淡出了人们的视线[34]。

然而从那时起，工程师们几乎并未放弃过 LINC。在有了先进的数字信号处理能力后又激励了许多人想借用这一强大的计算能力来克服 LINC 的某些不足。由此形成的 LINC 的通用的方框图显示在图 20.25 中。很遗憾，信号处理只满足了我们的部分需要。尤其是合成器的设计，它在很大程度上依然是无功而返。

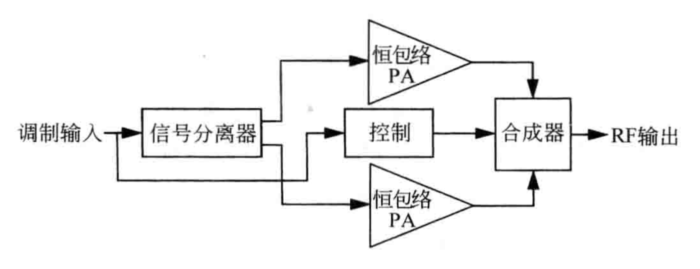


图 20.25　通用 LINC 系统方框图

极坐标反馈

由于通常需要同时校正在任何一个信号路径中的相位和幅值非线性，看来合乎逻辑的是对相位和幅值分别构成一个反馈环，如我们在讨论预畸变时所得到的启发那样。极坐标反馈（polar feedback）环可直接实现这一想法[35]。它常常与 EER（针对幅值分量）技术结合起来使用，并附加相位检测器及移相器，见图 20.26。

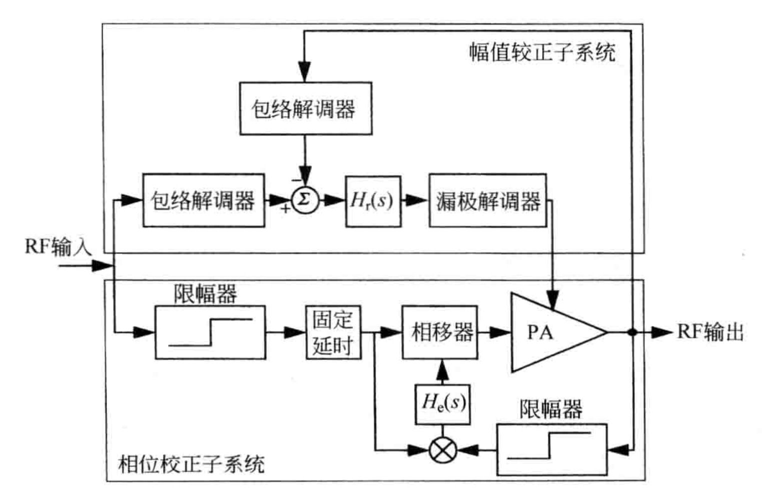


图 20.26　采用极坐标反馈的线性化功率放大器的例子

[34] 真的十分遗憾。

[35] V. Petrovic, W. Goslin, “Polar Loop Transmitter”（极坐标环路发射器），*Electronics Letters*, v. 15, 1979, pp. 1706-1712.



从图中很容易看出有两个控制环。幅值控制环对输出包络和输入包络进行比较，它们的差驱动一个漏极调制器，这与在 EER 中完全一样。增益函数 $H_\tau(s)$控制幅值反馈环路的动态特性。如果它包含有积分操作，那么稳态幅值误差就可以被驱动至零。由于大部分环路传输增益可以在基带（而不是在 RF）通过增益模块 $H_\tau(s)$得到，因此基本上可以在很大程度上抑制非线性。

相位控制环检查经过限幅的输入和输出信号间的相位差。实际上必须有限幅器，因为大多数相位检测器对幅值和相位都很敏感（或者要求具有某一最小幅值以便能够工作）。包括限幅器也很方便，因为基本的 EER 结构总需要一个。此时相位误差信号驱动一个与 PA 输入相串联的移相器并对它进行相应的调节，其动态特性仍然由一个增益模块控制，但现在的传输函数是 $H\theta(s)$。

一个重要的考虑是要保证幅值和相位校正在时间上相一致。然而由于相位和幅值控制子系统一般是用非常不同的元件实现的，因此不能保证它们的延时（或任何其他的相关特性）一定匹配。例如，幅值控制环的带宽与漏极调制器的带宽有关。如前所述，常常采用一个开关模式的调制器来保持高效率。但带宽很宽又将要求在调制放大器中有待别高的开关频率，其结果是：想用当前的工艺实现超出 lMHz 很多的带宽是非常困难的。

一般来说，相移环远未受到这么多的约束，因此它有宽得多的带宽和相应较小的延时。因而通常有必要插入一个固定的延时补偿以确保通过这两条控制路径的延时互相匹配。（从原理上说，移项器可以承担延时补偿，但此举在大多数情况下会不必要地使移项器的设计复杂化。）

由于极坐标反馈要求这两个类型非常不同的控制环必须具有延时匹配，因此想要始终一致地获得高性能显然很困难。另一个复杂性来源于 AM 至 PM 的转换，这种转换使这两个环路之间产生了互相耦合从而降低了稳定性。稳定性的问题又与幅值和 AM 至 PM 转换间的关系交织在一起。因此，极坐标反馈仍然是一个研究热点。然而现已证明，为了获得必要的匹配（更不必说要在电源电压、工艺和温度变化的情况下始终保持一致）以及为在输入全带宽及动态范围上确保稳定所带来的困难非常大，以至足以阻碍在移动通信最感兴趣的频率和带宽上实现大规模产品化。

注意，极坐标反馈所依据的一个重要概念是：把一个 RF 信号分解成两个正交分量，然后分别围绕每一个分量构造一个闭合的反馈环。因为极坐标的幅值和相位变量仅代表一种可能的选择，那么也许值得考虑另一种选择。

笛卡儿坐标反馈

一个信号的极坐标表示与直角坐标表示是等价的，因此可以不把信号分解成幅值和相位分量，而是把它分解成比如同相位的 I 和正交相位的 Q 分量。这个正交（Cartesian）表示有一些优良特性，因而正交反馈已受到相当程度的关注[36]。图 20.27 所示的方框图表明，与极坐标环路不同，正交反馈环路由两个电路上完全相同的路径构成。

---

[36] V Petrovic and C. N. Smith, "The Design ofVHF Polar Loop Transmitters"（VHF 极坐标发射器的设计）, *IEE Comms. 82 Conference*, 1982. pp. 148-155。还可参见 D. Cox, "Linear Amplification by Sampling Techniques: A New Application for Delta Coders"（采样技术的线性放大：一个 Delta 编码器的新应用）, *IEEE Trans Communications*, August 1975, pp. 793-798.

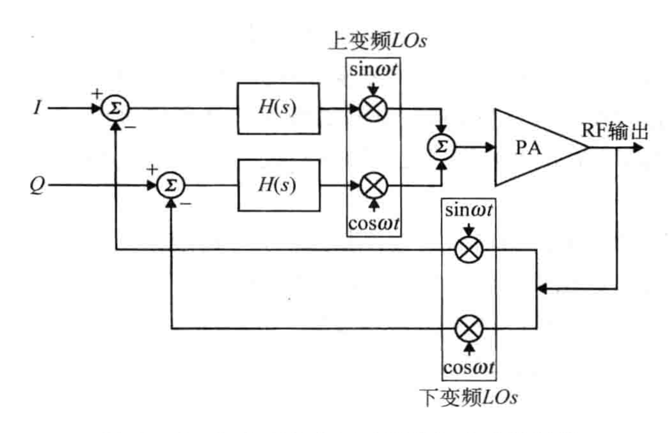


图 20.27 采用直角坐标反馈线性化的发送器

在这个结构中输出经过一对正交下变换。基带信号 I 和 Q 与它们调制后的相应信号进行比较。[37] 两个基带误差信号分别被计算、放大和反变换成 RF，并最终在 PA 级的输入端相加。大部分环路增益是在基带从 $H(s)$ 中而不是在 RF 中获得的，这大大简化了环路设计。

这个结构中两个反馈路径完全相同的事实表明，直角坐标反馈不存在极坐标反馈中棘手的匹配问题。但这里仍存在着困难的设计问题，这些问题又一次阻碍了该结构的广泛应用。

最严重的问题是在这两个环路之间缺少严格的正交性。只有保证正交性，这两个环路才能独立地工作，从而使设计时要考虑的问题较少。如果两个环路互相耦合，那么动态行为会以非常复杂(这又是中了墨菲定律的先验)的方式变化。更糟的是，两个环路不正交的程度有可能随时间、温度和电压而变化——并且也可能随 RF 载波在某个具体范围中调谐而发生改变。这个问题并不是直角坐标反馈所独有的，它可能在任何具有多条反馈路径的系统(例如极坐标反馈环)中都要考虑。

为了评估这一问题在系统级的影响，考虑在上变换和下变换 LO 之间有一个相位失调 Φ(仍然假设每对 LO 由两个正交的信号构成)。与任何反馈环一样，可以通过切断环路，注入一个测试信号，然后观察返回的信号来计算环路传输。这样做可以得到

$$L_{\text{eff}}(s,\phi) = L_{\text{one}}(s)\cos\phi + \frac{[L_{\text{one}}(s)\sin\phi]^2}{1+L_{\text{one}}(s)\cos\phi} \tag{41}$$

式中，$L_{\text{one}}(s)$是围绕每个独立环路传输值[38]。

这个有效环路传输表达式可以帮助我们理解为什么直角坐标反馈环可能表现出“奇怪的特性”。根据相位失调的程度，总的环路传输值可以从单环(此时失调为零)一直到两个单环的级联(此时失调为 $\pi/2$)。与许多控制环一样，$H(s)$多半设计成包含一个积分功能(以便使静态误差驱动至零)。如果失调仍然为零，那么这一选择不会有任何问题。然而当失调

[37] 假设我们或者已经完成了正交下变换以获得 I 和 Q 信号，或者事实上正实现我们已用数字技术产生的基带信号的上变换。因此该图说明的是一个发射器或一个放大器。

[38] J. Dawson and T. Lee,“Automatic Phase Alignment for a Fully Integrated CMOS Cartesian Feedback PA System”(用于全集成 CMOS 直角坐标功放系统的自动相位对准)，*ISSCC Digest of Technical Papers*, February 2003 .

增加时，$H(s)$会使环路传输增加两个积分极点，这将使相位裕量在最好情况下也为零。其他原因造成的任何负相移都会使相位裕量变为负值。

认识到这个机理是直角坐标反馈问题的一个重要来源，这还是近年来的事。解决的办法包括自动相位对准（以消除引起稳定性问题的来源）或仔细设计 $H(s)$使之能承受失调变化时环路动态特性的大范围变化[39]。慢滚降补偿是采用后一种策略时一种可能的 $H(s)$选择，它也可以与前一种策略结合起来使用。实现这些正确方法能使直角坐标反馈在极宽的带宽上对线性度有很大改善。

20.3.3　效率提升技术

在介绍了许多线性化方法之后，我们现在来集中讨论效率提升技术及结构。

自适应偏置

任何具有非零 DC 偏置电流的放大器的效率都会随 RF 输入功率的减小而降低（在这方面 A 类放大器比其他类放大器更差）。有许多 RF 的功率放大器（PA），例如那些用在蜂窝电话中的 PA，它们在相当长的一段时间内工作在低于最大输出功率的情况下，因此其平均效率极差。为了在低功率电平时提高效率，一个长期沿用的技术是采用自适应偏置（adaptive bias）策略[40]。根据放大器的即时要求动态地变化偏置电流和电源电压可以显著减少效率的降低（至少在原理上如此）。

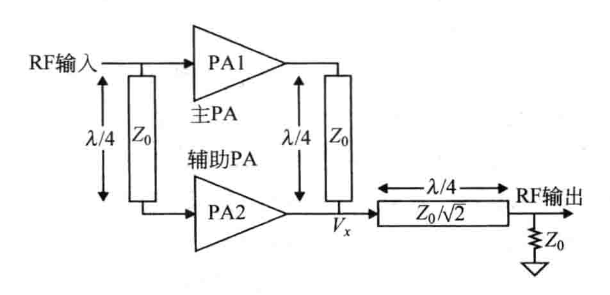


图 20.28　Doherty 放大器

在高调制值时，PA 级可以工作在相对高的电源电压下（相应地有较高的栅偏置）。在低调制值时，漏极电源和栅偏置电压也随之降低。如所期望的那样，这个策略大大削弱了效率与信号幅值之间的关系。由于灵活的低成本数字控制电路的出现，现在已能比以往更容易地实现自适应偏置。

可控制的漏极电源本质上就是 EER 系统中的漏极调制放大器，因而具有相同的设计挑战。自适应偏置总起来说还存在另外一个问题，这就是差不多同时改变如此多的重要参数是不利于线性性。然而，自适应偏置提供了一个额外的自由度可以用来实现效率和线性度之间一系列永无止境的折中选择。

Doherty 和 Terman-Woodyard 组合放大器

另一种效率提升技术是采用多个放大器，每个放大器负责整个功率范围内某个子范围

---

[39] Dawson，Lee. 出处同上。

[40] F. E. Terman 和 F. A. Everest，“Dynamic Grid Bias Modulation”（动态栅极偏置法调制），*Radio*，July 1936，p. 22.

上的放大作用。只采用能提供所希望输出功率的最少数目的放大器可以降低不必要的功耗。实际上我们实现的是相当于涡轮增压器的电子电路。这种想法的最早实现采用了两个放大器，它是由 Doherty 提出的[41]，见图 20.28。

放大器 PAl 和 PA2 分别是主放大器和辅助放大器。辅助放大器设计成在低幅值输入时截止。假设 PA2 的输出端在这种模式时为开路，那么很容易推导出从 PA1 输出端看到的阻抗为 $2Z_0$[42]。

在达到某一预定的阈值时，辅助放大器启动并开始向输出端提供功率。输入送至 PA2 的延时为 $\lambda/4$，它与 PA1 输出连至 PA2 输出间的延时 $\lambda/4$ 相匹配。因此，由 PA2 引起的 V_x 部分与 PA1 引起的 V_x 部分具有相同的相位。当 PA2 工作时 V_x 较大，这个事实表明从主放大器的输出延时线上看到的阻抗增大了，即 PA2 自举了 PA1。当 V_x 处增加的阻抗经 $\lambda/4$ 线映射回 PA2 的输出时，此处看到的阻抗减小了。这个 PA2 负载电阻的减小又使该放大器提供的功率加大。当两个放大器都提供各自的最大功率时，每个放大器看到的负载阻抗都为 Z_0 并且给输出提供相同的功率。

稍加思考就可以认识到，该组合放大器与 B 类推挽放大器有着相同的重要特性：它的一半功率由相应的一半电路提供。事实上，当两个放大器的分配与 Doherty 的最初设计一样时，极限峰值效率完全一样，约为 78%。从理论上说，平均效率要比这个值的一半稍低一些。

关于怎样选择辅助放大器开始给输出端提供功率的最佳阈值问题基本上可以通过考察该放大器将要处理的信号的包络概率密度函数(PDF)来回答。特别要注意应从调制方法日趋复杂的观点来考虑。例如，一个 16QAM 的信号理论上表现为峰值/平均值之比为 17dB。因此，一个平均输出功率为 16dBm(40 mW)的 16QAM 信号的偶然的峰值可能高达 33dBm(2W)。如果设计一个通常的 2W 放大器而它工作在 40 mW 的平均功率上，那么它事实上肯定具有极低的平均效率。类似于 Doherty 这样的技术似乎很适于这样的调制，因为一个高效低功率的主放大器将在大部分时间内承担这一任务，而辅助放大器只在断断续续地处理较少发生的高功率峰值时才起作用。我们看到当 PDF 在功率较低时密度较大，这意味着应当降低阈值。然而若遇到 PDF 在功率较高时密度较大的某种其他调制，那么我们就会希望提高这一阈值。可以采用许多种方法来实现任意的功率分配比率，其中包括调整耦合阻抗以及使这两个放大器工作在不同的电源电压[43]。

把功率范围划分成两个以上有可能进一步改善效率。虽然影响功率组合的负载结构的复杂度会迅速提高，但理论上的效率提升仍可能很显著。例如使平均效率加倍并不是不可能的[44]。

遗憾的是，Doherty 放大器与 B 类放大器并不只是在效率计算方面类似。

类似于跨零失真的问题也同样折磨着 Doherty 放大器。也许效率的获得是以失真为代

---

[41] W. H. Doherty, "A New High-Efficiency Power Amplifier for Modulated Waves"（一个高效率的调制波功率放大器），*Proc. IRE*, v. 24, 1936, pp. 1163-1182.

[42] 可以提醒读者：1/4 波导线可以用 CLC 的 π 网络来近似，如果这样做是更为恰当与方便的话。见 2.9.2 节。

[43] M. Iwamoto *et al.*, "An Extended Doherty Amplifier with High Efficiency over a Wide Power Range"（具有在宽的功率范围内高效率的扩展 Doherty 放大器），*IEEE MTT-S Digest*, May 2001, pp. 931-934.

[44] F. H. Raab, "Efficiency of Doherty RF Power Amplifier Systems"（Doherty 射频功放系统的效率），*IEEE Trans. Broadcast*, v. 33, September 1987, pp. 77-83.

价的，而我们必须做出很大的工程努力(例如把 Doherty 放大器嵌入在一个反馈环中)来同时抑制在幅值和相位域中非线性。与在错位相位系统中一样，在一个放大器输出端看到的阻抗与另一个放大器的输出有关。因此会有许多机会因预想不到的相互作用而出现许多不正确的行为。

对 Doherty 放大器的一个拓展是 Terman 和 Woodyard[45] 提出的调制器-放大器组合。它与 Doherty 放大器的类似之处是也采用了两个放大器(由经 $\lambda/4$ 延时的 RF 载波信号驱动)以及具有相同的输出合成器。但二者的差别是：它通过将同相位的调制信号同时注入到两个放大器的栅电路来实现调制功能。由于此时调制是器件传输特性中固有的非线性的结果，因此获得的调制结果不可能没有失真。但如果通过在包络信号上形成一个闭合的反馈环，那么就能够以较低的效率代价获得线性度较大的改善。

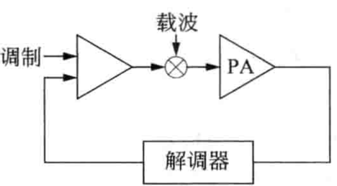


图 20.29　用来改进调制线性度的负反馈技术

最后，对所有这些方法，负反馈都可以被用来减少失真。为了缓解对反馈回路的增益带宽积的要求，可以对输出信号取样、解调，然后用已解调的信号形成环路。见图 20.29。这种结构常常根据所采用的解调的方式来区分。例如，如果解调器由一对为正交载波信号驱动的混频器组成，则功率放大器有时被称为通过笛卡儿反馈来实现线性化。

20.3.4　脉宽调制

另一个取得近乎线性调制的技术是采用脉宽调制(PWM)。采用这种技术的放大器有时称为 S 类放大器，虽然该名称并不是被普遍采用的。

这种放大器不是通过驱动幅值的变化来完成调制的。而是，这种调制通过控制恒定幅值驱动脉冲的占空比来完成的。这些脉冲经过滤波后，其输出功率就正比于输入的占空比，因此在原理上可以达到高效率线性工作的目的。

尽管 PWM 在较低的频率下(例如对于开关功率变换器直至兆赫范围中偏低的频率)工作得很好，但它在蜂窝电话千兆赫的载波频率下却完全没有用。其理由并不非常深奥。例如，考虑在 1GHz 的载波时达到 10∶1 范围调制的问题。由于半周期为 500 ps，所以调制到最大值的 10％要求产生 50 ps 的脉冲。即使我们能产生这样的脉冲(非常困难)，开关也多半不能完全导通，这有可能造成较大的功耗。因此在高频时，从本质上讲 PWM 放大器在输出功率的大动态范围内工作是没有希望的。换一个说法就是开关(以及它的驱动电路)必须比非 PWM 的放大器快 n 倍，这里，n 是所希望的动态范围。因此一旦载波频率超过 10MHz 左右，采用脉宽调制会变得越来越困难。

20.3.5　其他技术

采用串联提高增益或功率

级联是一种增加增益及功率电平的如此显而易见的方法，以至于仅仅提到这个方法就

[45] F. E. Terman and J. R. Woodyard"A High-Efficiency Grid-Modulated Amplifier,"(动态栅极偏置法调制)，*Proc. IRE*，v. 26，1938. pp. 929-945.

会被人蔑视。但在级联 PA 过程中存在着许多细节，值得我们不顾挨骂来提供一个简短讨论。

在放大级级联时，功率电平一般都逐级递增。如果前面几级消耗的功率足够小，那么可以仔细地设计这些放大级，把目标集中在线性度上，而把获得高效率的问题推迟到后面一级（或几级）去考虑。因此实际上最前面几级可以采用 A 类放大器实现，而最后一级才改为比如说 B 类或 C 类放大级。

当对放大级（例如 C 类）级联进行漏极调制时，漏极调制水平应当与功率水平成比例，使对级联放大级中每一级的驱动也按比例变化。如果不是这样的话，那就有过驱动某一级或几级的风险，从而出现与过载相关的问题，例如，从峰值恢复较慢、过度的 AM 至 PM 的转换以及差的线性度等。

最后，放大级的级联总有引起不稳定的风险，特别是当各级是调谐放大器的时候。这种风险对 A 类放大器最大，为了缓解风险可以采用一系列对低电平放大器行之有效的相同的一些通用技术。在比较棘手的情况下，当所有其他方法都无效时，衰减器可以提供解救办法。顾名思义，这种方法（如果可以将其夸大地这样称呼的话）是将一个电阻放在该电路的某处以丢掉增益和 Q 值。使这个电阻与栅极串联或跨接在栅-源两端是特别有效的手段。

采用注入锁定提高增益

在 16.2 节中讨论了有关注入锁定的问题，在这里只简述一下。我们应该简单地接受这样一个事实，即经常有可能把一个振荡器的相位锁定在注入到该放大器电路中某一合适点处信号的相位上，前提是倘若能满足一定条件的话（例如注入信号的频率足够接近振荡器的自由振荡频率，注入信号的幅值限制在一定的范围内等）。事实上，不希望有的注入锁定（也许由通过衬底耦合的信号引起）是 RF 集成电路中一个非常实际的问题。与其他寄生现象一样，事实上在一种情形下不希望有的某种效应有可能在另一种情形下被转变成有利因素。

虽然将放大器级联是一种提高增益的显而易见的方法，但 CMOS 器件较低的固有增益往往意味着必须有比在其他工艺（例如双极型工艺）中更多的放大级。这样做除了增加复杂性之外，所消耗的功率多半也会加大。为了避免这些问题，可以考虑构建一个振荡器，并以某种方式去影响它的相位或频率，这常常是很有利的。实际上，振荡器就是一个提供 RF 输出信号而不需要任何 RF 输入的放大器，因此“增益”是无穷大的。影响一个信号要比产生它更容易。由于影响锁定所要求的功率可以很小，所以表现出的增益就可以很大。

注入锁定作为一种放大技术通常限于恒包络调制，因为顾名思义，输入信号主要影响相位。从原理上讲它也可以提供幅值调制（例如通过改变偏置控制振荡幅值），但 AM 至 PM 的转换通常会很严重，因此这种组合只有在幅值调制深度较浅时才可行。

另一种可能是将注入锁定和错位相位技术结合起来产生幅值调制。这种组合在理论上表现为具有高增益和高效率，但是一个额外的问题是这又提出了许多细节的设计问题。这个结论的实验验证留给读者作为练习完成。

通过合成来提高功率

功率合成器可以替代或者与阻抗变换器一起来提供比用其他方法可以实际达到的更高的输出功率。一种常用的合成器是在第 7 章介绍的 Wilkinson 功率合成器，从理论上说，当

工作在匹配负载时它可以把功率没有损耗地合并起来[46]，见图 20.30。因此 Wilkinson 合成器可以把几个低功率放大器组合起来以提供大的输出功率。

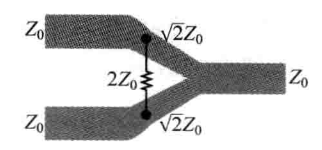


图 20.30　Wilkinson 功率合成器

如果输出功率需要超过 2∶1 的提升，那么就可以在一种称为合作合成器(corporate combiner)的结构中采用几级合成器，它的简图表示在图 20.31 中。这些技术通常在分立元件实现时都能成功，但在集成电路实现时却存在许多困难。首先，合成器各段传输线的长度为 1/4 波长，因此低千兆赫频率范围的功率放大器的合成器所占的面积会不切实际地大。举一个具体例子，1GHz 时一条片上 1/4 波长的线就长约 4cm。另一个问题是片上传输线的损耗。同样，1dB 的衰减就相当于 21% 的功率损失，但就连保持这样的一个个低损耗都有困难。而且如果 Wilkinson 合成器的终端不是理想地被终断，它不是无损的。任何不理想的终断所引起的反射都为桥接电阻所吸收，其后果是效率的损失。

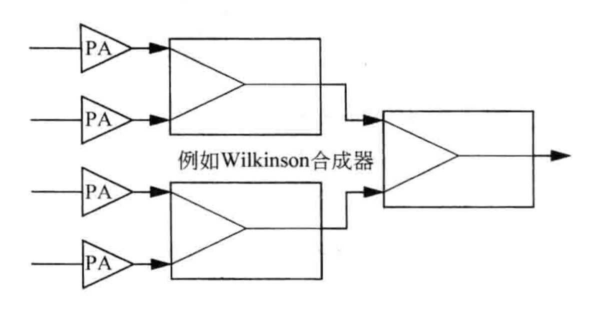


图 20.31　组合功率合成器

多级功放的另一种用法是并行地提供输入，然后把它们的输出串联起来。电压的提升使不切实际的阻抗变换比的要求被减小。该设想的一个非常简单的实现是建立一个差分功率放大器。在理想情况下，差分输出电压的摆幅是单端放大级的两倍，因此在一定的电源电压下可得到 4 倍的输出功率，并缓解对额外的阻抗变换的需求。可以通过平衡-不平衡变换器(balun)将差分输出转换为单端对地输出。

拓展这个概念的一个特别巧妙的结构是分布式有源变换器(distributed active transformer，DAT)[47]，这个名字的使用不很得当，因为该放大器并不是像传输线那样的分布式系统，比如不需要任何分布参数去描述它。更确切地说，这个结构的名字来源于它把总的功率分摊到许多器件中，它是"分而治之、各个击破"的具体体现。

假设需要使电压摆幅为原来的两倍以上以实现功率目标(或为了实现任何额外的阻抗变换比)，可以通过把两个差分放大器的输出加在一起以达到 4 倍的摆幅(因此使功率提高到 16 倍)。这种输出相加可以很容易地实现，即由各差分级驱动变压器的各个初级，然后只需把变压器的次级串联起来。图 20.32 说明了这个原理，它将其扩大到了 4 个差分级，理论

[46] E. J. Wilkinson，"N-Way Hybrid Power Combiner"，*IRE Transactions MTT*，1960，采用不对称结构还有可能提供不同的功率分配因子。

[47] I. Aoki *et al.*，"Distributed Active Transformer—A New Power-Combining and Impedance-Transformation Technique"(分布式有源变压器——一种新的功率合成与阻抗变换技术)，*IEEE Trans. Microwave Theory and Tech.*，v. 50，January 2002，pp. 316-331.

上可使电压摆幅（相对于单个器件的摆幅）提高 8 倍，这相当于使功率提升 64 倍（为简单起见，电路图或版图中没有表示出产生谐振负载所需要的电容）。

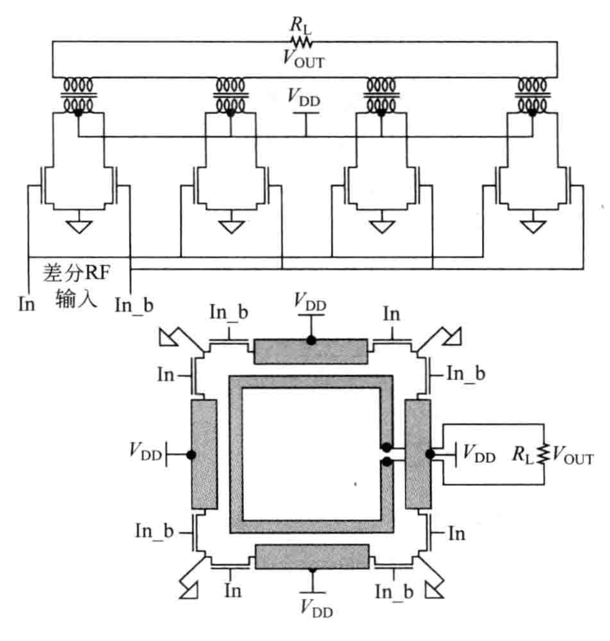


图 20.32 分布式有源变压器的示意图及简化了的版图（引自 Aoki 等，2002 年）

在图中所示的简化了的版图中，每个带中心抽头的漏极负载是用耦合线实现的片上变压器的初级。这里的次级是只有一匝的正方形电感，它的每一边和相应的中心抽头初级相耦合。由于四条边串联连接，因此它们产生的电压正如所希望的那样相加，于是在负载的两端产生了一个升高的输出电压。这里同时实现的阻抗变换意味着在次级中的电流是初级的 $1/N$，因此它可以采用比初级更窄的导线，如在图中用相对线宽（涂灰色部分）所表示的那样。

当将其一般化到有 N 个输出变压器的 N 个差分对时，可以看到，电压最大可提升为 $2N$ 倍（这里同样是相对于单个晶体管的电压而言），相应的功率可提升为 $4N^2$ 倍。虽然实际电路中的损耗无疑会使性能比这些最大极限值要低，但 DAT 仍不失为一种可行的选择。作为一个比较的参考值，这个概念最初实现时采用 0.35μm CMOS 工艺，在 2.4GHz 时可提供 2.2W 的饱和输出功率，其漏极效率为 35%（功率增加效率，power-added efficiency，为 31%）。输入和输出阻抗为 50Ω 时的增益约为 8.5dB。

20.3.6 性能标准

在复杂的调制技术出现之前，采用我们已介绍的指标来评价发射链的设计基本上是足够：这些指标包括饱和输出功率、三阶交调、1dB 压缩点、功率增加效率及类似的指标。例如，工程师们可迅速估计出为满足一定失真度所需要的功率回退水平，而且这常常足以建立起能工作很好的系统。对带外发射有比较宽松的法定限制曾一度反映了对这种尚不完善的最新技术的宽容。但这种情形已经由于采用比较先进的调制方法不断努力以减少宝贵的频谱浪费而发生改变。这种先进的调制方法增加了的复杂度也随之提高了对某些缺点的敏感

程度，因此有必要更为严格地标明和控制性能。

相邻信道功率比（adjacent channel power ratio，ACPR）是这类性能指标的一个例子。由于采用传统的双音（two-tone）测试来正确预测复杂数字调制（如 CDMA 系统）发射器所产生的干扰非常困难，因此提出了采用 ACPR 指标的方法，即通过采用有代表性的调制方式，然后直接测量对应于临近频道的频率的带外功率来表征干扰的可能影响[48]，见图 20.33。

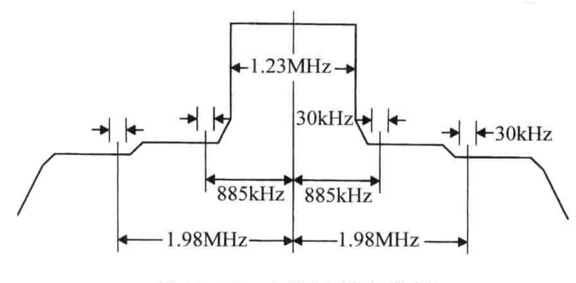


图 20.33　ACPR 规定举例

尽管在所有情形中将 ACPR 与 IP3 定量地联系起来是不可能的，但在典型的情况下 ACPR 仍能以非常类似于 IP3 的方式通过功率回退技术得到改善。也就是说，对于 1dB 的功率回退，可以期望 ACPR 有 2dB 的改善，同时带外边裙（out-of-band skirt）将做整体移动。对于以三阶线性度占主要地位的情形，这个结论常常（但不是总是）能成立。

测量 ACPR 的方法并不十分标准化，因此为了正确解释所报告的测量值的含义，必须知道这些值所对应的测量方法。举一个例子，IS-95CDMA 蜂窝式电话的 ACPR 在 885kHz 偏移频率处测量时应当比－42dBc 好[49]。一个细微的差别是有些技术测量功率密度积分的比，而其他技术是测量功率密度本身的比。其他差别包括积分带宽的选择。例如，我们可以在如图所示的 30kHz 的带宽上对 885kHz 偏移处的功率密度进行积分，然后除以在中心波束（central lobe）1.23MHz 带宽上功率密度的积分。严格来说，这个比值必须是－42dBc 或更好的值。

在其他（普遍得多的）测量中，功率密度在以两个测量频率为中心的 30kHz 的带宽上积分，然后用一个校正系数来将测量得到的比外推到与用前面方法所做的测量相对应。也就是说，在一定的假设条件下，这两个原始比值相差一个校正系数：

$$\Delta \text{ACPR} = 10\log \frac{1.23\text{MHz}}{30\text{kHz}} \approx 16.13\text{dB} \tag{42}$$

因此，用第二种方法测量得到的 ACPR 必须大约减去 16.1dB 才能与第一种方法得到的测量结果相对应[50]。第二种方法假设在载波中心 30kHz 范围内的平均功率密度与在 1.23MHz 频带其余部分的相同，但此时必须注意 IS-95 的信号一般会在这一频带上显示出 2dB 左右的波动。

---

[48] 然而，将双音测量与 ACPR 联系起来的解析方法依然能提供有用的设计内在了解。对这个途径的一个全面的讨论，见 Q. Wu *et al.*，"Linear and RF Power Amplifier Design of CDMA Signals: A Spectrum Analysis Approach"（CDMA 信号的线性与射频功率放大器设计：采用频谱分析的方法），*Microwave Journal*，*December* 1998，pp. 22-40。

[49] 严格说来，IS-95 定义了空中的界面，IS-97 规定了基站的性能，而 IS-98 则是指移动终端的指标。

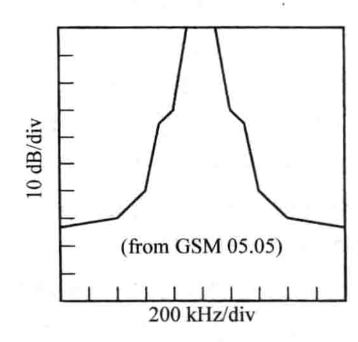
[50] 有关测量方法和对数据解释的更详细信息可参见"Testing CDMA Base Station Amplifiers"（测试 CDMA 基站放大器），Agilent Applications Note AN 1307。

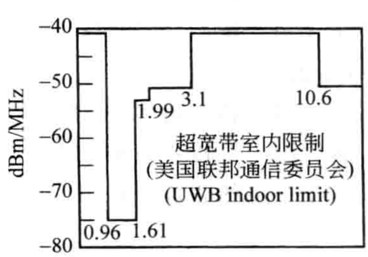
另一个在测量的细节之处是与用来求值 ACPR 所用的信号特性。对于 CDMA 系统来说，调制的本质是一个“类噪声”。作为一个结果，往往会倾向于采用经恰当的有限带宽的噪声作为 ACPR 的测试信号。但我们应当了解，噪声和类噪声是两件不同的事，其差异很像食品(food)和类食品(foodlike)之间的差别。ACPR 是失真的度量，因此它对平均功率的大小以及包络的细节(如峰值-平均值的比)很敏感。这些又与用来产生调制的码集(code set)有关。具有相同平均功率但不同的类噪声波形有可能使一个给定的放大器显示出相当不同的 ACPR。因激励不同从而使这些值在 15dB 的范围上变化的情况并不少见。

与表征失真效应的新方法一起发生的是出现了新的术语。听似像生物学的术语的频谱增生(spectral regrowth)是指由失真引起的频谱增宽。由于随功率水平的提高以及当信号向前通过发射器的各级时失真会加大，因而在分配发射链中各元件允许的失真指标时必须考虑这种频谱增宽的现象。因此，为了使整个发射器满足(例如)-42dBc 的 ACPR 指标，必须仔细设计合适的 PA 级，使它在最坏情况下的 ACPR 至少要比严格要求的指标好几个(例如 3 个)分贝，以满足整个发射器对这一 ACPR 指标的要求。

ACPR 的哲理可以用“作个个好邻居”来概括。指标的选择是本着这样的原则，即希望在符合这些指标时能够保证一个发射器与非目标接收器在其接收范围内之间的干扰最小。但一般来说，只在几个离散的频率处说明带外功率可能是不够的，例如就 ACPR 指标本身来说，并不能排除较强的窄带发射。在这些情形中也许必须说明的不是 ACPR 而是频谱屏蔽(spectral mask)。顾名思义，频谱屏蔽定义了对发射限制的连续区域。三个典型的例子表示在图 20.34 中。一个是 GSM(引自 05.05 版本的标准)，另一个是户内超宽带(ultra wideband, UWB)系统(如 2003 年 2 月 FCC Report and Order 中所定义的)，而第三个是 802.11b 无线局域网(LAN)。

特别是 UWB 屏蔽以其复杂度最为令人注意，在 0.96GHz 和 1.61GHz 之间的陷波主要是为了防止与全球定位系统(global positioning system, GPS)之间的干扰。这一屏蔽对 3.1GHz 至 10.6GHz 的限制最小，因此这段频谱常称为 UWB 频谱配置。同时应注意，屏蔽是用功率谱密度来表示的。由于在 UWB 系统中不存在载波，因此我们熟知的单位如“dBc”及“dBc/Hz”不再适用。





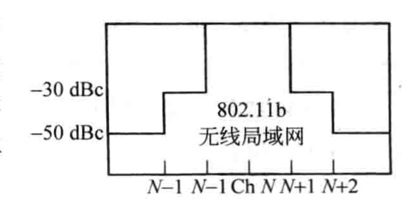


图 20.34 发射频谱屏蔽举例

通过符合这些屏蔽标准来满足“好邻居”的要求是必要的，但却是不够的。还必须保证发射器产生的调制能使所希望的接收器成功地解调。误差向量幅值(error-vector magnitude, EVM)特别适合于用来对许多数字调制发射器中的不希望的行为进行量化。误差向量这个概念很自然地可应用到采用某种(例如 QAM)向量调制的系统中。

误差向量幅值就是误差向量的长度，如图 20.35 所示。每个符号(symbol)或片码(chip)都有自身的误差向量。对于 802.11b 标准的 11 Mb/s 的 WLAN，EVM 定义为 1000 个片码的均方根(rms)值，并且它必须小于一个非常宽松的值，35%。而 802.11a 标准工作在 54Mb/s(标准中说明的最大数据传输率)时，允许的 EVM 为 5.6%——这比 802.11b 的规定要严格得多。

在测量 EVM 时，仪器会提供一个看上去像图 20.36 所示的图。在理想情况下星座图表现为完好的 8×8 方点阵，图中模糊不清的就是这些点。对每一个模糊斑点的归一后的均方根值就是 EVM。在所示的具体这个例子中，EVM 大约为 2%——这完全满足了指标要求。

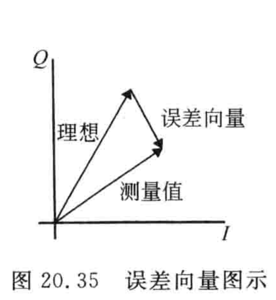


图 20.35　误差向量图示

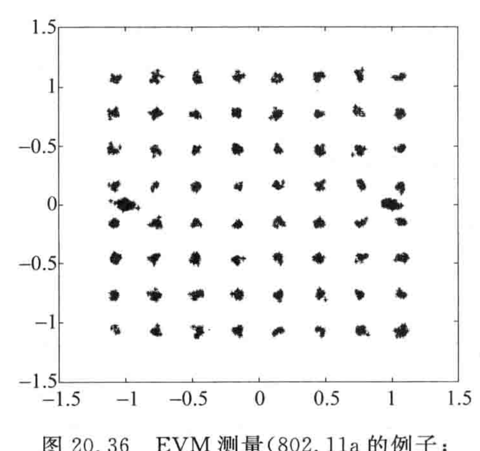


图 20.36　EVM 测量(802.11a 的例子；所示为 64-QAM 星座图)

20.4　其他设计考虑

20.4.1　附加功率效率

在前面的例子中，集电极效率用来表征功率放大器。然而，集电极效率的定义只涉及到 RF 输出功率及 DC 输入功率，因此一个没有任何功率增益的功率放大器也可以具有很高的效率。由此提出了另一种效率的衡量指标来产生一个把功率增益考虑在内的品质因子。附加功率效率(power-added efficiency，PAE)就是在集电极效率公式中用输出和输入功率之间的差来代替 RF 输出功率：

$$\mathrm{PAE} \stackrel{\mathrm{def}}{=} \frac{P_{\mathrm{out}} - P_{\mathrm{in}}}{P_{\mathrm{DC}}} \tag{43}$$

显然，附加功率效率将总是小于集电极效率。

20.4.2　功率放大器的不稳定性

任何种类的放大器在负载和信号源阻抗的某种组合下都有可能出现不稳定，功率放大器也不例外。一个极为重要的问题来自于集电极至基极的耦合(或者漏至栅的耦合)。正如在第 12 章中所指出的那样，这种耦合有可能使输入阻抗具有一个负的实数部分。在小信号

放大器中，这个问题可以通过前面介绍过的各种单向化技术来减轻或完全消除。可惜的是这些技巧对于功率放大器一般来说并不合适，因为对于高效率的要求将阻止使电源电压裕量减小的任何技术(如共栅技术)的应用。一般来说，这个问题通常通过迫使输入阻抗减小(例如简单地采用输入端口间接一个电阻)以使反馈不太显著来解决。遗憾的是，这个办法具有降低增益的副作用。一般来说，MOSFET——由于它们本身的输入阻抗较大——都会比双极型器件显示出更大程度的稳定性问题。不管怎么说，由于反馈电容的存在通常都会有一个严重的稳定性—增益之间互换的问题，而且很自然，需要进行仔细的版图设计以避免由于输入和输出连线间的并列布置而加大内在的器件反馈电容。

不稳定性甚至会出现在开关型的放大器中。因为实际的驱动波斜率是有限的，使得开关器件有可能在线性区保持足够长时间来产生振荡。即使在振荡还未发生时，输入阻抗(举个例子)可能经历相当大的变化。因此在设计前级电路时必须考虑这种可能性以避免那里不稳定的发生。

20.4.3 击穿现象

MOS 器件

在所有的设计实例中都要求采用向下的阻抗变换以把所希望的功率传送到输出负载中。显然，如果可以允许较高的电源电压，那么阻抗变换比率可以减小，因此读者也许有理由问为什么我们不可以要求采用一个较高的电压。其理由是器件具有有限的击穿电压，而且随着 IC 工艺缩小到越来越小的尺寸，击穿电压也趋于减小。因此如果希望传送给负载某一固定数量的功率，那么当器件缩小时就要求增大阻抗的变换比率。

在 MOS 器件中，对在功率放大器中允许应用的电压有 4 个主要的限制，它们是漏(或源)二极管的齐纳击穿、漏源间的穿通、与时间有关的电介质击穿(TDDB)以及栅氧的破坏。

漏区和源区都进行重掺杂以降低它们的电阻率，因此它们与衬底之间形成的二极管具有较低的击穿电压，对于 0.5μm 工艺来说其典型值在 10～12V 的数量级上。漏源间的穿通类似于在双极型器件中的基极穿通，并且发生在当漏极电压很高以至于足以引起漏极附近的耗尽区一直延伸到源极时，从而在事实上消除了沟道，于是电流的流动就不再受栅极电压的控制。

与时间有关的电介质击穿是高能量载流子造成栅氧破坏的结果。在现代短沟器件中典型高电场的作用下，载流子(主要是电子)有可能被加速到很大的能量以至于足以使它们在栅氧中形成陷阱。因此任何在陷阱中被捕获的电荷就会使器件发生阈值偏移。在 NMOS 晶体管中，阈值将增加，使得在给定栅电压下得到的电流减小，而在 PMOS 器件中情况恰好相反。与时间有关的电介质击穿有累积的效果，因此它限制了器件的寿命。一般来说，TDDB 规则设计的目标是在 10 年后驱动电流的减小不超过 10%。

双极型器件

双极型晶体管不存在任何栅氧的损坏，但其结的击穿和基极穿通是对电源电压的重要限制。当电场足够高从而引起空穴-电子对的显著产生且成倍增长时，集电极-基极结就可能发生雪崩击穿。在很好设计的器件中，这一机理会导致更为严重的限制，尽管作为高截止频率(f_T)器件特征的极薄基区常常使基区穿通也成为非常重要的限制因素。

另一个更为微妙的能给双极型器件带来严重破坏的问题是无法减小的终端电感，这个

电感与大的 di/dt 一起发生作用。当关断器件时，会有显著的基极电流沿反方向流动直到基区的电荷被抽净为止。当基极电荷被抽净时，基极电流会突然停止流动，于是大的 Ldi/dt 就可能在基极与发射极之间引起很大的反向电压尖峰。回想一下基极-发射极结具有较低的反向击穿电压（例如 6～7V，虽然有些功率器件具有较大的值），并且由击穿造成的破坏取决于（载流子的）能量且有累积的效应。特别是 β 将减小（并且器件噪声也更严重）。因此增益就会减小，还可能会引起不正确的偏置，并且输出频谱中失真乘积项可能增加以及最低噪声底限将不断恶化。因此在对功率放大器进行模拟时，重要的是要特别关注这一效应并且在必要时采取一些补救措施[51]。其中可选择的措施包括在器件两端连接一个箝位二极管（也许可以与晶体管本身集成在一起以减少在箝位二极管和输出晶体管之间的电感），再者就是通过改进的版图设计或更好的驱动控制来减小。

在 MOS 实现中有可能（但很少）发生类似的现象。在关断期间当栅驱动减小时，一旦栅电压降至阈值电压以下，栅电容会突然减小。同样，Ldi/dt 尖峰电压也可以大到足以使器件被破坏的程度。

20.4.4　热失控

另一个问题关系到热效应。为了实现高功率工作，人们通常采用并联器件。在双极型中，集电极电流为常数时基极-发射极电压的温度系数约为 $-2\mathrm{mV/^\circ C}$。为此，当器件变热时，它要求较小的驱动即可保持所需要的集电极电流。所以对于一个固定的驱动，当温度升高时集电极电流会急剧增加。

现在考虑在几个并联的双极型器件中如果一个器件比其他器件更热一点时会发生什么情形。随着温度的升高，这个器件的集电极电流会增加，因此它会变得更热并产生更大的电流。如果这一热电正反馈回路的传输率超过 1，那么这个正反馈回路就可能失控，使器件迅速损坏。为了解决这个问题，在每个晶体管的发射极端加上某个小的电阻负反馈极为有用。采用这种方法时，如果在任何一个器件中集电极电流有增加的趋势，那么它的基极-发射极电压就会降低，于是抵消了负的 T_C（温度系数），避免了热失控。许多制造商都把这样的负反馈（常常称为镇流）集成到器件结构中，所以不必在外部增加任何元件。即便如此，由于这一正反馈的原因，在高功率放大器中会经常见到温度的差别在 10℃或以上的情况。

热失控在 MOS 实现中通常不是一个问题，因为随温度升高迁移率的下降使得在固定的栅-源驱动时漏极电流减小而不是增加。一个例外可能发生在当反馈控制用来使栅-源电压随温度上升以保持恒定驱动电流时。在这种情形下，器件的损耗随温度上升，从而有可能重新造成热失控的情形。

无论是双极型还是 MOS 功率放大器，最好都要包括某种形式的热保护以避免过载。幸运的是，在 IC 实现中采用片上温度计测量温度并设法相应地减少器件驱动是很容易做到的。

20.4.5　大信号的阻抗匹配

在功率放大器的大信号范围内驱动晶体管时，如果想要使阻抗与驱动源很好地匹配，这

---

[51] 再次要说明的是：重要的是要有一个基于实际物理结构的好模型。

会是一个严重的挑战。这种说法对双极型晶体管和场效应管都适用。比如，由于基极-发射极的结毕竟是一个二极管，所以输入阻抗是高度非线性的。双极型晶体管的制造商注意到了这一困难，所以常常针对一个具体的功率大小和频率来说明输入阻抗。然而，由于一般没有任何可靠的指南来说明它如何随功率大小或其他工作条件变化，并且由于甚至没有保证可以找到一组条件使阻抗匹配得以实现，所以设计者的设计选择十分有限[52]。解决这个问题的传统方法是用一个小数值电阻连接在基极与发射极之间以消除非线性的问题。如果这个电阻足够小，那么它的电阻值就在输入阻抗中占支配地位。在高功率设计中，这个电阻可以小至1Ω或更小。同样，以上说法对双极型晶体管和场效应管都适用。

在有些情况下，不能随便加一个并联电阻，但允许提高电路的复杂度，这时可以通过使用正交耦合器来获得很好的输入输出匹配。其中3dB兰格(Lange)耦合器和分支线耦合器均可用在图20.37所示的结构上。

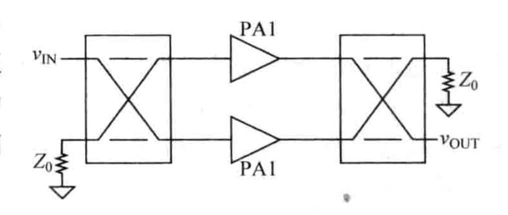


图20.37 用来产生输入输出匹配的正交混合耦合器

两个功率放大器因输入失配或输出失配而产生的反射将由图中的终断电阻所吸收，因此输入输出端口在理论上可看做是理想匹配。每个功率放大器可以工作在有针对性的失配条件下，比如，根据需要可以使其输出最大功率或者拥有最好的稳定性。与此同时，该正交耦合器也是功率合成器，使得每级功率放大器只需提供一半的总输出功率。然而，需要牢记的是标称的失配越严重，对电阻的功率耗散容忍度要求也越高。

最后，对频谱纯度性的符合标准的要求在采用简单输出结构(如在这些例子中采用单个振荡回路)时不会总是得到满足的。通常必须串联上附加的滤波部分以保证可接受的低失真。可惜的是，每个滤波器都不可避免地带来了某些损耗。就这方面而言，重要的是记住只是1dB的衰减就代表着非常大的21%的损耗，因此要求十分小心地注意处理好所有的损耗来源以保持较高的效率。

20.4.6 功率放大器的负载拉特性

至今我们考虑的所有例子都假定负载是一个纯电阻。遗憾的是，现实的负载很少是纯电阻性的。特别是天线对功率放大器来说从来不表现出其额定的负载，因为它们的阻抗受不可控制的变量如与其他物体的靠近程度(例如在蜂窝电话应用中的人头)的影响。

为了探讨一个变化负载阻抗对功率放大器的影响，我们可以系统地改变负载阻抗中的实部和虚部，然后在阻抗平面(或等效地在史密斯图)上画出恒输出功率的等值线，这些等值线合起来称为负载拉(load-pull)图。

为了得到负载拉图的近似形状，可以仍然假设输出晶体管的工作情况在摆动的全过程中可以作为一个理想的可控电流源。下面的推导引自S. L. Cripps的经典论文[53]，他首先

[52] 这里且不去管器件制造商们有帮助的应用说明，这些说明通常都具有乐观的标题，如"系统的方法使得C类放大器的设计变得轻而易举"等。

[53] "A Theory for the Prediction of GaAs FET Load-Pull Power Contours"(一个有关GaAs场效应晶体管负载拉功率等值线预测的理论)，*IEEE MTT-S Digest*，1983，pp. 221-223.

将其应用到 GaAs 功率放大器中。

假设放大器工作在 A 类模式，于是负载电阻与电源电压及峰值漏极电流的关系如下：

$$R_{opt} \stackrel{def}{=} \frac{2V_{CC}}{I_{C,pk}} \tag{44}$$

相应的输出功率为

$$P_{opt} \stackrel{def}{=} \left[\frac{1}{2}I_{C,pk}\right]^2 R_{opt} \tag{45}$$

现在，如果负载阻抗的数值小于这个电阻值，那么输出功率就由电流所限制，因而在这个电流限制情况下传送到负载的功率就是

$$P_L = \left[\frac{1}{2}I_{C,pk}\right]^2 R_L \tag{46}$$

式中，R_L 是负载阻抗的电阻部分。

峰值集电极电压是峰值电流和负载阻抗幅值的乘积：

$$V_{pk} = I_{C,pk}\sqrt{R_L^2 + X_L^2} \tag{47}$$

用公式(44)替换上式中的峰值集电极电流得到

$$V_{pk} = \frac{2V_{CC}}{R_{opt}} \cdot \sqrt{R_L^2 + X_L^2} \tag{48}$$

为了保持线性工作，V_{pk}的值必须不超过 $2V_{CC}$。这个要求限制了电抗负载部分的值：

$$|X_L|^2 \leqslant (R_{opt}^2 - R_L^2) \tag{49}$$

可以将以上顺序排列的公式解释如下：对于负载阻抗幅值小于 R_{opt} 的负载，峰值输出电流限制了功率；在阻抗平面上恒输出功率的等值线就是恒定电阻为 R_L 的线，直到达到公式(49)所给出的电抗极限为止。

如果负载阻抗的幅值超过了 R_{opt}，那么所传送的功率由电源电压限制。在该电压摆幅限制的情况下，比较方便的是考虑负载的导纳而不是负载阻抗，因此输出的功率为

$$P_L = [V_{CC}/2]^2 G_L \tag{50}$$

式中 G_L 是输出负载导纳中的电导部分。采用与前面类似的方法可以计算集电极电流为

$$i_c = 2V_{CC}\sqrt{G_L^2 + B_L^2} \tag{51}$$

式中 B_L 是输出负载导纳中的电纳部分。公式(51)中集电极电流的最大值为

$$i_{C,pk} = 2V_{CC}G_{opt} \tag{52}$$

把公式(52)代入公式(51)并求解这个不等式得到

$$|B_L|^2 \leqslant (G_{opt}^2 - G_L^2) \tag{53}$$

我们对上面这些公式的解释是：当负载阻抗的幅数大于 R_{opt} 时，恒功率等值线是电导 G_L 为常数时的线，直到电纳的值达到公式(49)给出的值为止。这两个阻抗区域的等值线一起构成了负载拉图。

负载拉图对于估计 PA 对负载变化的敏感程度、确定最优的工作条件以及有可能揭示隐藏的弱点是非常有价值的。有经验的 PA 设计者常常具有通过观察负载拉图来诊断出许多问题的能力。

20.4.7　负载拉等值线举例

为了说明这个步骤，让我们对一个假想的 A 类放大器构筑负载拉图，其中峰值电压为

6.6V，峰值电流为1.65A，因而R_{opt}为4Ω。为了求出使可以输出的功率为在最优设计值偏差1dB处的所有负载导纳（或阻抗）的轨迹，应该首先计算出从4Ω偏离1dB时相当于约3.2Ω和5.0Ω。前一个值用于电流限制的情况，而后一个值用于电压摆幅限制的情况。

在电流限制情况下，沿着3.2Ω常数电阻线直至最大允许的电抗幅值约为2.6Ω时为止，而在摆幅限制情况下沿着0.2S的常数电导线直至最大允许的电纳幅值为0.15 S时为止。

习惯上不是把等值线画在阻抗和导纳平面上，而是画成史密斯图的形式。由于阻抗或导纳平面上的圆在史密斯图中仍保持为圆（直线被认为是半径为无穷大的圆），所以这些等值线中有限长度的直线段就变成史密斯图中的圆弧。相应的负载拉图显示在图20.38中，这里归一至5Ω和0.2 S（而不是50Ω和0.02 S）以使等值线较大，足以被清楚地看出。

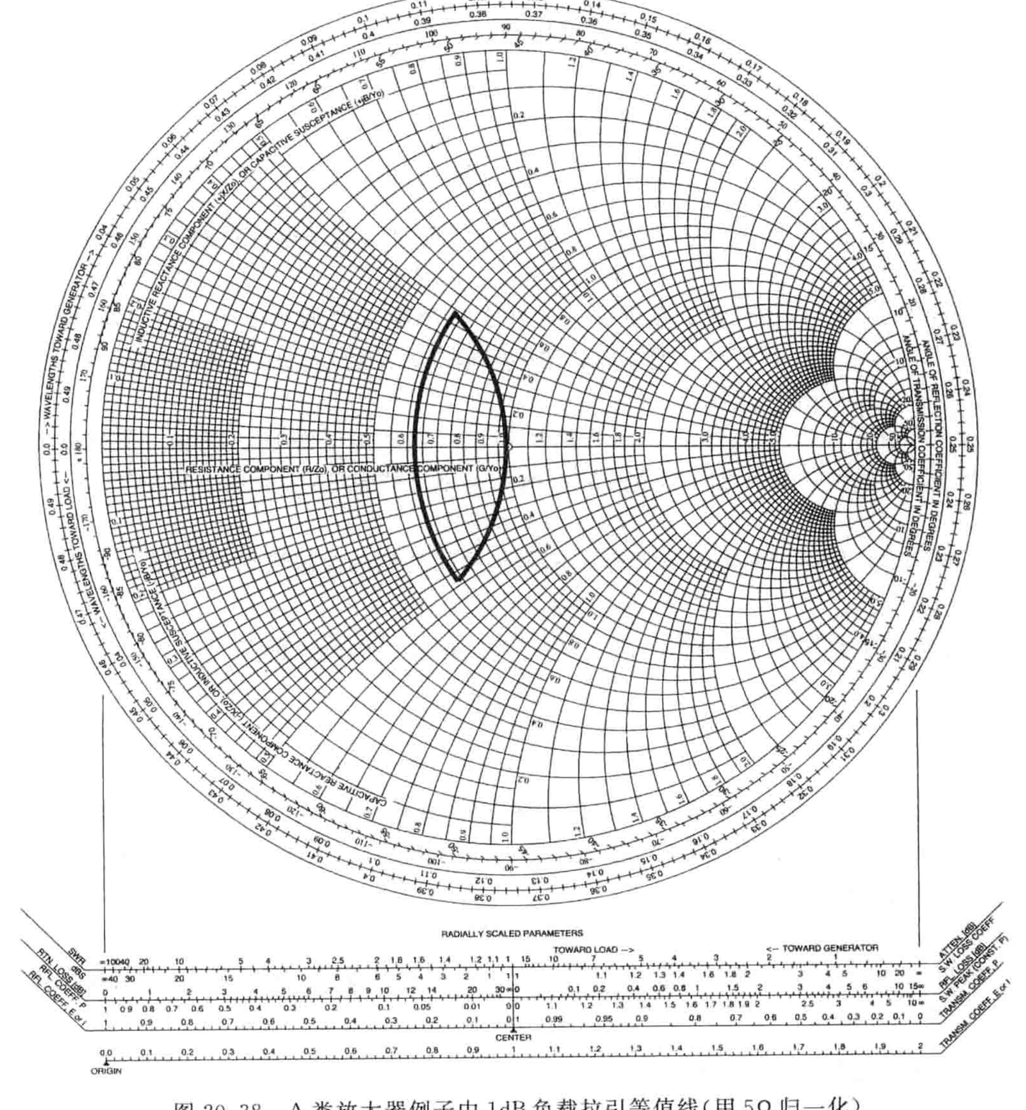


图20.38　A类放大器例子中1dB负载拉引等值线（用5Ω归一化）

传送到负载的功率因此对处于两个圆交叠部分内的所有负载阻抗，都将在最大值的 1dB 范围内。这两个圆中的一个是常数电阻(它的值比最优负载电阻小 1dB)，而另一个是常数电导(它的值比最优负载电导小 1dB)。从以上描述中注意到不需要计算电抗或者电纳的极限数值，这两个圆的交叠部分已自动地利用图解完成了这个计算。因此建立理论的负载拉图要比进行细节推导的过程容易得多。

这里应当强调一下，前面的推导假定了晶体管的工作如一个理想的、没有寄生参数的可控电流源。器件和封装的寄生参数连同外部的负载阻抗一起构成了这一负载拉图的总等效负载。然而在建立实际的负载拉图时只知道外加的阻抗，因此基于外部阻抗值的负载拉等值线一般将需要对没有寄生参数的情形进行转换、旋转与变形。

实例注解

功率放大器的负载牵引表征依赖于可调节经校准了的调谐器。实部和虚部可以顺利地调节到需要的值，在此过程的每一步中增益和输出功率都要来测试得到。可以想象，完成一个完整的测试将是一件非常麻烦的事情。自动测试仪器可以减少人力的劳动，但是这些仪器非常昂贵。

目前存在两种商用的负载拉系统：一个基于步进电机驱动的机械调谐器。它们一般体积大而且速度比较慢，但是具有很好的精度和可重复性；另外一种方法是使用电子开关元件(比如说 PIN 二极管或者其他半导体开关)来实现调谐。这些仪器比传统步进电机的调谐器更紧凑更快。然而，对于全电子化的方案，与调谐器相关的寄生效应十分严重，需要用户对校准和补偿花多得多的注意。

20.5 总结

我们已经看到 A 类功率放大器能提供好的线性度，但是其功率处理能力较差(归一化值 0.125)且效率低(理想最大值 50%)。B 类功率放大器通过减少晶体管在一个周期内有源工作的时间来提高效率(最好情况～78.5%)，同时也保持着线性调制的潜力。

C 类功率放大器提供的效率接近于 100%，但是归一化的功率处理能力和功率增益同时接近于 0。同样的，它牺牲了线性处理能力以提高效率。另外，双极性 C 类功率放大器实际上不满足很多在我们推导过程引入的假设，因此很难来设计和构造，在这点上，MOS 管和真空管相对会简单些。

基于开关理论的放大器尚未能提供线性调制方式的能力，但是它们(理论上)在非零的功率处理能力上能提供 100%的效率。尽管这个完美的效率并不能实现，至少这种限制不是这种结构的固有问题。

D 类功率放大器提供归一化的功率处理能力接近于 0.16，但是它们遭受潜在的大“过零”损耗，这是由于非无穷大的开关速度而造成的。E 类功率放大器通过开启过渡解决了耗散问题，但是会导致对开关产生非常大的电应力。E 类和 F 类功率放大器使用电抗负载以改造电压和电流波形，以降低开关耗耗，这表明了这个方法可以有效地提效率。

我们也考察了许多提高线性度和效率的办法，而且以考虑一些测量表征方法来结束本章。

第21章 天线

21.1 引言

在这本书的前面部分，我们就已经注意到了这样一个重要认识，即传统的集总电路理论是从近似宇宙的实际行为而来的(具体地说，将麦克斯韦方程某些项设定为 0，这样就有效地将光速认为是无穷大)。那几个被过分视为神圣的基尔霍夫[①]"定律"实际上完全不是定律，它们是作了简单近似的结果，因此最终它们是要失效的[②]。电路理论的集总参数描述——其中将基本元件分为阻抗、电容和电感——在元件尺寸相对波长较小时是允许的。尽管严格证明这个长度标准已经超出了本书的主要讨论实际问题的宗旨，给出一个简短的可行性的论证还是许可且已经足够了。

我们可以将下面这个观察接受为公理，即当光沿着长度为 Δl 的电路部分中传播的延时 T_D 只占我们感兴趣的最小时间段 $T_{\min}$ 的一小部分时，那么光速的有限性是可以忽略的，即要求

$$\left(T_D = \frac{\Delta l}{v}\right) \ll T_{\min} = \frac{1}{f_{\max}} \tag{1}$$

其中，v 是传播速度，$f_{\max}$ 是我们所感兴趣的最高频率。这个不等式可以整理为

$$\Delta l \ll \frac{v}{f_{\max}} = \lambda_{\min} \tag{2}$$

波长 $\lambda_{\min}$ 是所感兴趣的最高频率信号的波长。

只要电路元件相对于相关的波长"足够短"，传统的电路分析因此是有效的。你可能会试图来争辩这一"小元件"的限制并不是一个严格实际的限制，因为总可以将大结构分解为合适的小元件，从而每一个元件都可以利用集总近似来准确描述。但是，这个方法的问题在于隐含地假设了所有网络中的能量是被限制在了电路元件所在的空间。在这一章中，通过考虑电磁能量辐射的可能性来摒弃这种假设。这么做的话，必须确定出现严重辐射的条件。可以看到理论上对于任何长度的导体都可以产生辐射，但是对于那些尺寸至少为波长的相当一部分的结构会更加容易。理解对长度的依赖就解释了为什么总是可以在低频下忽略辐射，以及为什么大部分经典的天线都是那么大。

就如关于滤波器的讨论，天线这个题目是太过广泛而无法在简单的一个章节给予全面

---

① 可以注意到，Kirchhoff 这个名字中有两个 h，两个 f，其发音作"keerk off"而不是"kirtch off"。

② 不承认这个事实是层出不穷的伪命题的根源，其中不少在各种互联网论坛上进行辩论(如"证明了那个物理定律是错的"，等等，这是被那些经常出错而自己又意识不到的人干的事)。

的处理[3]。这里,我们的主要目的是:(a)培养一种在某些课本中的严格数学处理里很少能提供的(或者说很难从中提取的)直觉洞察力;(b)补充在许多"如何做"中通常给出的过于简略的解释。因为这一章目的是为了补充而不是替代已有的描述,因此要注意有时会牺牲一些严格的推导来培养设计所需的洞察力。实际上,因为在讨论过程中有这么多只涉及定性与表层的描述,读者在理解时要注意其中还存在着更深层次的内涵。比如说,在本章中,你不会看到任何含有磁矢势的积分。

除了这种定性叙述带给读者的轻松感觉外,这个做法的最能切身感受到的结果推导出对天线的简单解析电路模型,以及体会到为什么在实际应用中会有如此之多的不同的天线结构。

尽管本书的重点是在平面电路上,我们在一开始先介绍对(电)偶极子天线的研究,这不仅是因为它有广泛的应用,也因为对它的分析能够揭示所有经典的天线带的共性问题。对偶极子局限性的清晰的理解能够帮助解释为什么某些对结构的改动,比如加入"容性帽"(capacity hat)或者负载线圈,能够大大提高短的偶极子的辐射特性。将会看到,这同样的理解也揭示了在归一化长度、效率和可达到的带宽之间的关系,这种关系使我们想起了在许多放大器中可以发现的增益-带宽的折中。

描述偶极子的那些方程也直接导致了对磁环路天线的描述,因为它们是对偶的,环路天线就是一个磁偶极子天线。为了与我们的平面电路讨论角度相一致,这一章花了相当一部分时间来研究微带线贴片天线,这种天线在最近几年变得非常的普遍,因为它很容易通过用制作印刷电路板同样的工艺来低成本、大量地制造。

因为我们最关心的问题是为天线建立等效电路模型,因此其他的一些重要特性(比如辐射模式、方向性和增益)在这里都遗憾地被忽略了。有兴趣的读者可以参考脚注 3 的参考文献。

21.2 坡印廷(Poynting)定理、能量与导线

为了形成一个统一的观点来解释一根金属导线什么时候是一根导线,什么时候是一个天线,非常关键的一点就是要放弃心理上那个想象——电就像流体一样流过像管道那样的导线——这是在本科教育或以前被有意识地灌输给学生。替而代之的,我们需要理解的是理想的导线严格地说不会携带任何电磁能量[4]。许多(或许大多数)学生和工程师,更不用说是门外汉,认为这个命题是有争议的。但是,导线不会携带能量这个命题是正确的,并且很容易来证明。

为了证明该命题,从普通的低频电路理论中的功率公式开始:

$$P = \frac{1}{2}\mathrm{Re}\{VI^*\} \tag{3}$$

[3] 有关这个课题的三本极好的教科书是:Stutzman and Thiele 的 *Antenna Theory and Design*(天线理论与设计)(Wiley, New York, 1998),Kraus 的经典教材 *Antenna*(天线)(McGraw-Hill, New York, 1988),及 Constantine A. Balanis 的 *Antenna Theory*(天线理论,第二版)(Wiley, New York, 1996)。对于业余无线电工作有用的有关天线构建的实际信息可以在由美国无线电继电器联盟(American Radio Relay League, ARRL)出版的 *The ARRL Antenna Handbook*(ARRL 天线手册)及其他繁多的书中找到。

[4] Heaviside 或许是第一个明确说出这个想法的人。

简而言之，传输实际功率需要有电压、电流和它们之间恰当的相位关系（公式（3）中的星号表示复共轭）。如果 V 或 I 中任意一个为零，那么就不会有功率传输到负载上去。进一步说，即使两者都不为零，完全正交的相位（90°相位）关系仍然可能导致无法传递实际功率。

与低频电路相应的电磁场论中对同样思想的理论表达式是坡印廷（Poynting）定理，其指出与电磁波相关联的（实）功率与电场及磁场的矢量叉积成正比[⑤]：

$$P = \frac{1}{2}\mathrm{Re}\{\boldsymbol{E} \times \boldsymbol{H}^*\} \tag{4}$$

为了传输实际功率，必须有 $\boldsymbol{E}$、$\boldsymbol{H}$，以及它们之间合适的相位。如果 $\boldsymbol{E}$、$\boldsymbol{H}$ 中有一个为零，或者完全正交，那么没有功率会被传输。现在我们看到，在完美的导体中电场为零，因此根据坡印廷定理，没有实际能量会流过一根导线；如果有任何的能量流动，那它一定完全在导线以外的空间[⑥]。许多学生很轻松而正确地利用坡印廷定理解决了电磁场论中高深的研究生问题，但是却对这一特殊必然结果表示成文字时的理解上显得十分的困难，因为它似乎违背了常识和日常生活的经验（“我只有碰到了导线，才会被电击”）。

对这一看起来似非而是的命题的解释是导体引导电磁场能量的流向。这种回答似乎是在玩文字游戏，但它却是可以帮助我们建立对导线、天线、电缆、波导甚至是光纤的统一认识的深入理解。因此为了使这本书的讨论（也是使你的职业生涯）保持前后一贯性，读者要牢记导体是引导电磁能量，而不是电磁能量在其中流动的管道，如果没有这根导线，电磁能量会是弥散在整个空间的。有了这个概念，那么许多看起来不同的传递电磁能量的方式就可以简单地看作是一种方式的变种。

21.3 辐射的本质

不少的学生已经习惯了无奈地面对这样一个事实，电动力学那些方程——充斥着梯度、散度和旋度——是一种发明来算计与折磨不幸的本科生的。也许从一个教师的角度来看，这毫无疑问是电磁场理论最有价值的特性之一。

或许可以理解，由这一痛苦引起的精神上的折磨常常会使得学生忽略了更为重要的问题：准确地来说，什么是辐射？一根导线是怎么知道什么时候它是作为天线，而什么时候不是？天线的端口电特性是什么样子的？而这些端口电特性又是如何为附近的物体所影响的？谁发明了液体皂，为什么？[⑦]

让我们以熟悉的集总电路理论为例开始。为了不失普遍性，考虑用正弦信号源驱动一个纯电抗（比如说一个无损电容器或电感）。如果仔细观察电压电流特性，我们会发现它们

---

⑤ 说起来都叫人不相信，这个理论是由一个真实的人的姓命名的。这个人的名字是 John Henry Poynting，他在 1884 年在赫兹开始他的那些实验之前提出了这个定理。坡印廷矢量正好指在能量流动的方向上，这只是为数不少的令人吃惊的巧合的例子中的一个。

⑥ 当考虑实际的导体时，这个说法要稍加修正。此时，会出现一个小小的沿着导线的切线方向的电场，这个电场强度只要大到可以解释其中发生的欧姆损耗就可以了。

⑦ 电影 *The Sure Thing*（必然发生的事）（Embassy Films Associates，1985，导演 Rob Reiner）中的 John Cusack。

是精确地在时间上正交("ELI the ICE man"—ELI 这个冰(ICE)人,和所有那些类似的叫法[8])。信号源传输到纯电抗的平均功率为零,因为能量仅是在源与电抗之间来回流动。在一个 1/4 周期中,比如说,一定量的能量流入电抗,而在下一 1/4 周期中,所有这些能量又返回到信号源里。为了使传输的平均功率不为零,必须要有同相的电流和电压分量。在一个电抗两端(或与之串联)加入一个电阻将会对纯 90°相位关系引入一个相位偏移,使得能够产生一个同相的分量,从而也就产生了与此相应的功率耗散。

对于电磁场的功率流问题也包含完全一样的考虑。当电场与磁场完全时间正交时,不会有实际的功率流。如果定义辐射是一种对远处放置的负载的功率传输,缺乏实际功率流因此也就意味着没有辐射。我们早就知道(从集总电路理论)正交关系在标称的电抗中是占主要地位的,只要频率所对应的最短波长相对电路尺寸要大得多。比如,将连接到地的一根短导线看作是电感,而将悬浮在接地平面上的导体看作是电容。这些处理方法是可行的,因为围绕导体的电磁场变化十分"缓慢"。

就如那些集总电抗的例子一样,实际功率传输需要非 90°相位关系的电场和磁场。经典的天线通过光速的有限速度来产生这样额外的延时使得偏离正交相位[9]。

为了具体地理解光速的有限性是如何能够产生(实际上是使得发生)辐射,考虑一段有限长度的导体在一端被正弦电压源驱动。接近源处的电场和磁场可以很好地被近似为正交的。但是由于信号沿导体传播的时间不为零,导体另一端的电压(从而电场)相对于被驱动端的电压和电场有滞后。两端处的电流(以及它所对应的磁场)也相似地有偏移。因此远端的电场不再与源端的磁场成 90°的相位,从而有非零的平均功率从驱动源传递出去。从定性上出现相似性质的集总电路的类比如图 21.1 的网络所示,其中这个电容被两个同频率的正弦波发生器所驱动。

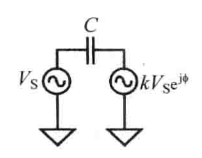


图 21.1 为两个同频率但不同步(即相位有差别)信号源驱动的电容

粗看一下这个特别的网络,可能会倾向于得到没有功耗的结论,因为电容是一个纯电抗元件。实际上,这也是博士生在资格考试时中最常见的给出的结论。让我们通过计算,比如说从左边信号源看过去的阻抗来直接检验这一命题的正确与否。通过电容的电流可以简单地由电容上的电压除以电容的阻抗来得到。因此,

$$Z_{eq} = \frac{V_S}{V_S(1-ke^{j\phi})sC} = \frac{1}{sC(1-ke^{j\phi})} = \frac{1}{j\omega C[1-k(\cos\phi + j\sin\phi)]} \tag{5}$$

其中常数 k 是一个任意的实数(这里并不是表示玻尔兹曼常数)。注意方括号中的因子当相位角 ϕ 是 0°或者 180°时为一纯实数。在这种条件下,阻抗的相位角为±90°,意味着损耗为零。能量仅是在一个 1/4 周期内被存储到电容里,而在下一个 1/4 周期还给了两个信号源。但是任何其他的相位角都会产生等效的从信号源端看上去的阻抗的一个实分量。尽管只存在一个纯电抗,损耗依然是可能的。电容当然仍不消耗功率,但是这里还有两个电压源需要考虑。非零平均功率在这两个同频率的源之间传递是可能的。也就是说一个源能够对另一

⑧ 假若这个俏皮语对你是生疏的话,这是一个帮助记住在电感与电容器间的阻抗相位关系的方法。"ELI"指的是,在电感里(L),电压(或电场,E)超前于电流(I);而在电容中,电流超前于电压。

⑨ 当然,这个与正交相位的偏离在电路中存在着真实电阻时也会发生。我们这里的注意力是在辐射上,所以将不再考虑这种带有损耗性的机理。

个做功。

类似的思想可以应用到辐射问题中。由于传播速度有限，那些通常仅在导体周围的空间里储存能量的电场与磁场分量，突然变为能够为远处的负载传递功率，这就是辐射。结果是，那个驱动导体的信号源一定会看到除了已存在的电抗以外的一个等效电阻。对这个结果一种想象方式是可以认为这一电阻和辐射是由于移动天线中一部分的电荷对天线中另一部分电荷所做的功而引起的[10]。与辐射相关的场其实是始终存在的（能量并不在导体内，而是在空间里），但是辐射只有在合适的相位关系时才发生。

从之前对辐射的描述，理解为什么天线的长度十分重要就不难了。如果导体（天线）非常短，那么时间延迟就很小，导致相对正交的偏离就很小。更准确地说，当导体长度相对于波长非常短时，天线阻抗的电阻分量就会相应地很小。相对于波长进行归一化是很有道理的，因为一个给定的长度会产生一个固定的延时，而这一延时反过头来又代表了当频率增加（即波长减小）时线性增长的相移。

现在已经推演出辐射是非纯正交的一个必然结果，让来看一下能否推导出辐射对于距离的依赖关系。回想到在自由空间中静态的孤立电荷的电场与距离的平方成反比，可能会倾向于说辐射也一定表现一个平方反比的规律[11]。为了对这一想法进行测验（还是尽可能少用到场论），假设有一个电磁能量的源（目前，完全没有必要给出更具体的条件）。考虑从源向外在两个像邻的同心球面上的能量流。如果存在辐射，那么通过两个球壳的总能量流必须相等；否则总能量会随距离增加或者减小，也就是说有能量的产生或消失[12]。因此可以得到：

$$\text{energy} = \bar{P}_1 A_1 = \bar{P}_2 A_2 \tag{6}$$

其中 $\bar{P}$ 是功率面密度，A 是面积。现在，由于一个球面面积是正比于半径的平方的，总能量守恒就意味着功率密度必须反比于距离的平方而衰减。在自由空间中，电磁场是相互成正比的。与坡印廷定理想联立，可以知道功率密度正比于场强的平方：

$$\bar{P} \propto |E|^2 \propto |H|^2 \propto 1/r^2 \tag{7}$$

因此，可以看到必存在电场或磁场的某一个分量随距离的一次方衰减，从而使辐射成为可能[13]。这一发现十分重要，因为如果只能依赖于具有平方反比空间依赖性的场（比如由一个孤立静态电荷产生的），那么长距离的通信会十分困难（因为呈 $1/r^4$ 的功率衰减会是一场灾难）。幸运的是，一个电动力学的奇迹产生了时变的电磁场分量，其在空间中的衰减大为减缓（再一次强调，这是在自由空间里）。正是这些辐射分量使得无线通信变成了现实。尽管还没有推出有关场的精确形式，但是毫无疑问已经从一些非常初步的论证推出了有关这些场的重要特点。

---

⑩ 已过世的美国加州理工学院的诺贝尔（物理学）奖获得者，Richard Feynman，以在所有现存的描述方式中最为简洁的方式来描述这个过程："如果你摇晃一个电子，就会发出光来"。也就是说，辐射不仅仅是由被加速的电荷所带的场与其他电荷的场交互作用的结果（那些其他电荷可以在天线内部，也可以在天线周围的媒介中），而且也可以来源于电荷与其本身产生的场交互作用。这就是说，能产生辐射的非正交的 $\boldsymbol{E}$ 场与 $\boldsymbol{H}$ 场并不需要来自一个结构的不同部分。

⑪ 因为不存在着绝对速度这样的事，可以预见到（根据相对论的基本原理考虑）辐射不会从电荷的一个匀速运动，必须要有加速，才会有辐射。

⑫ 或者是一个在自由空间在单调增加的能量存储，而这是与稳态的假设所不相容的。

⑬ 重要的是要记住这个结论是依赖于在自由空间中传播的假设。如果这个假设不成立（比如说，因为有损介质的存在），那么会得到其他不同的结论。

除了使我们将辐射与反比于距离的场联系起来，之前的讨论还告诉我们对于信号源来说，辐射的能量和耗散在电阻上的能量必然是不可分辨的。其结果是，我们可以看到，如前所述，辐射为天线的驱动点阻抗提供了电阻分量。

同时注意到至此的推导实际上给出了一个相当强的结论：除了反比于距离的空间依赖性，没有任何别的距离依赖可以与在自由空间发生的辐射相关。比如，如果场衰减得过快，那么空间中能量就会在两个相邻的球面间积累。如果场衰减过慢，那个空间就必须提供能量。由于这两个条件都不可能在稳态时满足，可以得到这样的结论：这样的场分量是不能支持辐射的。而那些其他的分量最多是代表了存储的（电抗性）能量，这些能量在源与周围环境之间来回流动。因此，它们的效应在天线的电路模型中是感性或者容性元件（取决于能量是主要存储在磁场还是电场中）来建模，我们很快会转向这些模型的发展。

在没有依赖于高深的数学的情况下，我们已尽可能多地得到了内在理解。现在在考虑天线的构建与建模的实际问题之前，先对天线的性能特征进行描述。

21.4　天线特性

天线性能的表征往往是相对于一个假象的各向同性的辐射器来进行的。正如它的名字所指的，各向同性的天线向各个方向均匀地发射能量。尽管实际并不存在各向同性的天线，但它为实际的天线提供了很有用的归一化参考。

实际的天线当然不会向每个方向辐射相同强度的能量。对于这种非各向同性的测量包括了方向性增益和定向性。为了理解这些测量，首先假设天线辐射的总功率为 P_{tot}，天线的等效各向同性功率密度（equivalent isotropic power density，EIPD）在距离 r 处即简单的将总辐射功率除以半径为 r 的球面表面积[14]。

$$\mathrm{EIPD} = \frac{P_{tot}}{4\pi r^2} \tag{8}$$

定向性增益定义为在某个特定方向上的实际辐射功率除以等效各向同性功率密度（EIPD）。定向度就是简单地为定向性增益的最大值。

前面的这些定义都是有关辐射的功率。因为在任何实际的天线中都存在着损耗，总的辐射功率会小于传输给天线端口的功率。为了计入损耗，可以引入一个电效率：

$$\eta = \frac{P_{rad}}{P_{tot}} \tag{9}$$

其中，P_{tot} 是提供给天线电气端口的总功率。

天线增益这个笼统的术语包括了定向性与低于 1 的电学效率的共同效应。与这些量的测量相关联，人们经常会遇到“dBi”这个单位。参考单位“i”指的是各向同性的辐射器的性能。因此，说一个天线增益为 3dBi，指的是那根天线在某一特定的方向上辐射出较一个各向同性辐射器高 3dB 的功率。因此，当与天线一起使用时，增益指的不是实际功率增益（这是很荒谬的，因为天线本身是无源的）而是相对于一个标准天线的增益。这个标准天线典型的就是一个各向同性辐射器，但有时也会是一个偶极子天线。

---

[14] 这个归一化通常是与一个单位半径的球面连在一起进行的，这样方程(8)中的 r^2 就不出现了。

令人迷惑的是常常在书中出现的全向天线(omnidirectional antenna)这个词。与人们想的相反,它并不是各向同性天线的别称。它常常被用来描述那些定向增益在某个平面内与角度无关的天线,但并不一定是对所有的平面都是如此。因此所有的各向同性天线都是全向的,但并不是所有的全方向天线都是各向同性的。

考虑到实际天线都是各向异性辐射的,一个单一的数,如方向性或者增益,只能提供关于天线性能的不完整的描述。波束宽度这一参数因此被用来补充这些量度,波束宽度描述功率密度在峰值功率的 3dB(或者其他某些标准值)以内的角度。

最后,许多天线被设计来将它们的能量集中在某一个方向上,但实际上做不到那么理想。它们的辐射图形可以表明存在着在其他方向的能量传输。对于这种旁瓣辐射的控制是许多天线设计的一个目标。

图 21.2 展示了具有代表性的辐射模式(这个例子是对一个传统的偶极子天线而言的,我们很快就会更具体地讨论这种天线)。所用的视角是从侧面看过去(即是一个立面图)。天线是一根直棍竖在圆形的中心,放置在一个占据了下半圆的导体地上。可以看到在沿着天线本身和地方向上完全没有辐射。辐射在这两个极限之间的某个角度达到最大。在这一例子中,最大点出现在仰角为 27°左右(接近光标位置处),天线的波束宽度(在-3dB 点处测量)为大约 44°。

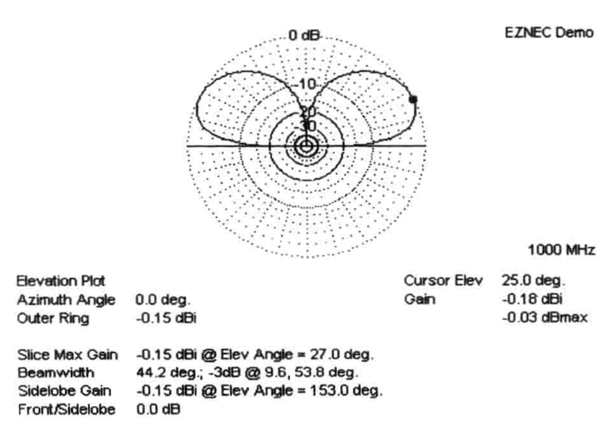


图 21.2 具有代表性的天线辐射图(采用 EZNEC 3.0 模拟器模拟；立面图)

看一看 3D 的辐射图像也是很有启发性的,因为它清楚地展现了全向辐射的本质(再一次强调,这个全向性只在一个给定的平面内存在),见图 21.3。

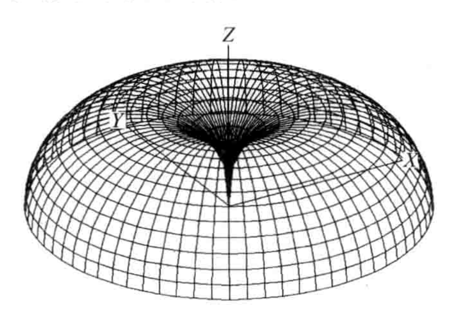


图 21.3 图 21.2 所示天线的三维辐射图

21.5　偶极子天线

最常见的天线毫无疑问是偶极子(包括电偶极子与磁偶极子)。可能你对图 21.4 中电偶极子相当熟悉。无数个看起来像“兔耳朵”的偶极子天线几十年来坐在电视机顶上(现在仍然这样),而同样是无数个的偶极子天线可以在现在的手机和汽车里看到(举个例来说,仅 2002 年就出售了超过 4 亿台手机)。那些低级的、无处不在的调幅收音机也采用了磁偶极子形式的天线。如我们将要看到的,偶极子天线的工作原理是直接根据所给出有关辐射的说明。

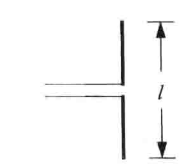


图 21.4　短的中心馈入的偶极子天线

21.5.1　辐射电阻

我们曾论证过,辐射必然会导致在输入阻抗中电阻分量的产生。这个辐射电阻特别重要,因为它决定了,比如说,能量是如何有效地从信号源被耦合到辐射的能量。同时,通过对偶原理,我们可以知道天线电路模型对发射与接收过程是相同的。

从最基本的原理出发来推导偶极子的辐射电阻是困难的,这是因为这种天线在这种推导之前已经使用了很长时间了。然而,非常有用的工程近似是容易被推得的,这只要通过简单地假设一个电流沿天线的分布。这个推导过程是由一个合理的假设所指引,即认为偶极子可以近似地被认为是一个终端开路的由两根线组成的传输线(的确是,这是一根被弯折的传输线,但是以后你会发现,如果你的要求过于严格,你会后悔的——即推不出一个简单的结论来)。然后,就得到了一个近似的正弦电流分布,其边界条件是在传输线的开路端电流接近于零[15]。

采用这一假设的电流分布,并且假设天线是由无穷薄的超导体组成的,可以推导得到下面这个短偶极子天线的辐射阻抗的近似表达式:

$$R_r \approx 20\pi^2 \ (l/\lambda)^2 \tag{10}$$

这一公式与数值模拟(基于同样的假设)的结果相当接近,直到 l/λ 的值接近约 0.3。在半波长时,近似公式预测辐射电阻大约在 50Ω 左右。在实际情况中,实际天线的阻抗在共振点时具有的典型实部值接近 73Ω。但是,当频率略高于共振频率时,阻抗接近 60Ω,这为 50Ω 传输线提供了很好的匹配(由一个匹配网络提供的有限的改进,几乎一定会被在任何实际实现时必然存在的损耗所抵消)。因此,近似公式(10)的实际有效范围在比其准确的预期范围更宽。

如果认为天线比它的实际长度略长,这个公式的准确度还可以提高一点。这一有效的扩展来源于这样一个事实:由于边缘场的作用,电流在天线的顶端并不完全为零。之后会看到,有时人们可以故意增强这一效应来改进(即增加)短偶极子的辐射电阻。

这个长度修正因子是关于半径与波长之比的有点复杂的(但依赖关系相当弱的)函数,并且在经典的偶极子天线中这个修正因子常常被近似为 5%。也就是说,实际长度应该被

[15] 因为存在着非零的边缘效应造成的电容,这个电流其实不完全为零。但这个电流为零的假设只引起了很小的误差,于是其后的推导就足够准确了。

乘以大约 1.05，再将乘积代入到公式(10)中去。

马可尼的主要发明之一是一种极富价值的偶极子天线的变种，在这个天线结构中，地面（或其他导体平面）中感生的镜像电荷有效地使得天线的长度加倍。对于这样一个竖在难以捉摸的完美的接地面上的单极天线（如图 21.5），辐射电阻对于式(10)给出的值会被整整加一倍[16]。

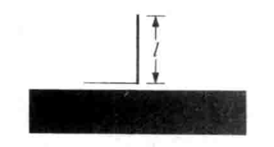


图 21.5 在接地面之上的短单极天线

有效长度的加倍使辐射电阻变为原来的 4 倍。但是只有竖直单极的物理部分（非镜像部分）才会产生辐射，从而又使 4 倍的辐射电阻减半（明白了么?）。因此，短单极天线的辐射电阻为

$$R_r \approx 40\pi^2 \ (l/\lambda)^2 \tag{11}$$

这个公式在(l/λ)值直至大约 0.15～0.20 时是相当准确的。相比较公式所给出的 25Ω，无限薄的超导体的四分之一波长单极子天线大约会有 36Ω 的阻抗。同样地，实际的单极子更有可能在共振点以上表现出最小驻波比(SWR)，而典型的长度修正也常常保持在约 5%的量级上。

由模拟得到的天线驻波比如图 21.6 所示，我们可以看到驻波比在 15%的总带宽里都低于 2∶1。光标所在的位置表示天线阻抗虚部变为零的频率位置。在这个频率下，这一例子的阻抗实部为 36Ω。最小的 SWR 出现在更高的频率，这时天线的电抗部分呈略感抗。虽然 SWR 从未达到 1∶1，但是它在很大的频率范围内已经足够低，以至于即使加入匹配网络对系统性能的改善也很有限（特别是考虑到无损的匹配网络是不存在的）。

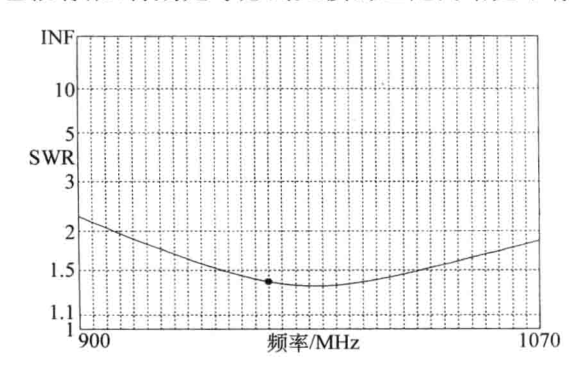


图 21.6 在 1GHz 附近的单极天线的电压驻波比(SWR)与频率的关系图（采用 EZNEC 3.0 模拟结果）

就像开路的传输线一样，单极子和偶极子都展现周期谐振的特性。辐射电阻的函数形式随着谐振模式而变化。对中心供电的偶极子，辐射电阻可以用如表 21.1 所示的十分近似的公式来计算[17]。

---

[16] 单极天线也经常被称为偶极子天线，这是因为就它们的工作原理及电流分布而言，并没有根本的差别。在本书中，这两个术语都被用到，读者可以从上下文中体会到确切的含义。

[17] 在前面已经引用过的 Stutzman and Thiele 一文（见脚注 3），p. 171。这些公式继续假设无穷薄的导体，因此它们给出的值不会精确地反映实际天线具有的值。同时，仔细的读者会注意到这些公式在从跨越同一长度的不同表达式得到的结果并不一致。比如说，在精确的四分之一波长情况下，跨越这个长度的两个公式给出的结果不相符，这会给读者造成困惑。但这对所有近似来讲，是不可避免的。

表 21.1　短的和中等长度的中心馈入偶极子天线的近似辐射电阻

归一化的导体长度，l/λ	R_{rad}
0～0.25	$20[\pi(l/\lambda)]^2$
0.25～0.5	$24.7[\pi(l/\lambda)]^{2.4}$
0.5～0.64	$11.1[\pi(l/\lambda)]^{4.17}$

表中公式经过修改也可以用在单极子上。只要记住镜像使得天线的有效长度加倍；在公式中插入这一加倍的长度（还要包含约 5%的长度修正）。然后，记住只有实际的天线（而不是它的镜像）才会辐射，使得辐射电阻减半。最后，公式适用于归一化长度的范围为表中的第一列中数值的一半。再一次说明，这里的长度是实际导体的长度，而不包括镜像。

最后，要知道我们很少会遇到完美的地平面，这一点也是十分重要的。如果试图利用导电薄片来产生这样的平面，那么它至少要有四分之一波长的半径。如果这样的薄片是不现实的（比如在低频率下，波长很长），几根那么长的导体棒也可以产生让令人满意的近似接地平面。

21.5.2　天线阻抗中的电抗分量

由于辐射带走了能量，它的效果可以用一个电阻作为损耗来模拟。但通常有些能量仍然保留在天线周围，在源与周围的空间之间来回流动。这种近场非辐射分量就代表了存储的能量，因此它会给端口阻抗贡献一个虚部分量。

为了得到非常近似的短天线的有效电抗（电感和电容）的表达式，我们再一次将偶极子天线比作开路的传输线。如果假设了 TEM 传输以及相对介电常数和磁导率均为 1，那么光速可以表示为

$$c = 1/\sqrt{LC} \tag{12}$$

其中 L 和 C 为每单位长度的电感和电容。现在，已经有了下面的方程（见第 6 章）来近似圆形截面的导线的电感：

$$L \approx \frac{\mu_0}{2\pi}\left[\ln\left(\frac{2l}{r}\right) - 0.75\right] \tag{13}$$

因此，单位长度的电容（F/m）可以十分近似地为：

$$C \approx \left\{c^2 \frac{\mu_0}{2\pi}\left[\ln\left(\frac{2l}{r}\right) - 0.75\right]\right\}^{-1} \approx \frac{2\pi\varepsilon_0}{\ln(2l/r) - 0.75} \approx \frac{5.56\times10^{-11}}{\ln(2l/r) - 0.75} \tag{14}$$

对于典型的尺寸，单位长度的电容可以非常近似地认为具有 10pF/m 的量级。比如，一根 10cm 长的规格 18 的导线（半径大约 1mm）根据公式有几乎精确的 1pF 电容。注意电感的增长是比线性增长要快些，而电容增长要比线性慢些。因此，单位长度的电容不是一个常数，但是 10pF/m（或者 10fF/mm）的估计值在许多粗略计算中能够适用得很好。就像与其他的一切事情一样，如果你需要更高的精度那你就需要做更多的工作。幸运的是，电磁场求解器可以处理这些更加具有挑战性的情况。

再一次把偶极子天线当作开路的传输线，我们期望短偶极子展现出主要为容性电抗，然后随着增长线长到共振（半波长）而转变为纯电阻，过了共振点后又转变为感抗。这种总体

趋势是周期性的，重复每一个波长，但是峰峰的电抗变化会慢慢消失因为随之增加的累积损耗[18]。也可以推断出单极子的驱动点阻抗也是每半波长地周期性变化的[19]。

最后说明一下，必须要重申所有这些方程都假设了天线是处在自由空间里，周围没有任何其他物体（除了单极子中的地平面）。实际天线的测量往往会与这些简单的公式所预测的有很严重的偏差，但主要是因为人们很少能够让实际的环境中所有物体都与天线保持很远。与天线距离小于几个波长的物体可以对天线的电抗和辐射电阻有非常大的影响。不十分严格地说，天线的电抗主要对周围的电介质物质非常敏感（如果天线的工作主要是基于电场）或者对磁介质物质非常敏感（如果天线主要依赖于磁场）。而实部则通常是对周围的有损介质最为敏感。

总结一下这一节的结果，简单的无损偶极子可以用图 21.7 的简单电路来建模。在这个模型中，发生器代表了收到的信号所产生的电压。对于短偶极子来说，这一电压就是电场强度与天线长度的乘积。当天线用来发射时，这个发生器被设定为 0（即短路），辐射功率可以通过传输给 R_{rad} 的功率来计算。任何损耗（比如由于趋肤效应产生）可以加入一个额外与辐射电阻串联的电阻来模拟。

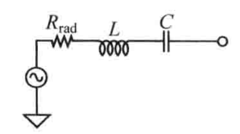


图 21.7　偶极子天线的电路模型（仅对一个模而言）

21.5.3　容性负载偶极子

我们已经看到短偶极子辐射电阻随归一化长度的平方变化。因此良好的辐射要求偶极子长度相对波长而言占相当的比例，否则辐射电阻会太小使得无法高效地将能量耦合到辐射中（或从中提取出来）。遗憾的是，随意增加天线的长度来满足这一要求并不总是可行的，特别是在低频率下（记住，1MHz 下自由空间的波长为 300m）有时主要的限制是来自机械工程的问题：即要支撑又高又细的天线带来的困难。

一种解决这个问题的方法是把天线弯过来（支撑一个长的水平物体总比高的竖直物体来得容易）。为了理解为什么这样是很有益的，回想一下这么一个现象：竖直偶极子天线的边缘场会使得天线的行为仿佛其物理长度比实际的长度要长。与边缘场相联系的电容会阻止电流在边界上变为零，从而增加天线的平均电流，进而增加辐射电阻。尽管这一效果通常很小，对长度的修正仅仅占普通偶极子长度 5%，但是边缘效应可以被有目的地利用来将短偶极子的长度显著提升[20]。在那些由于竖直方向的空间限制而不允许长偶极子天线应用中，人们可以利用容性负载——利用所谓的电容（容性）帽来增加在端口处的电流，同时也增加了沿着偶极子长度的平均电流。有各种各样的导体布局可以使用，包括用平面圆碟、球面和水平的线（后者常常在 L 和 T 形天线中使用）。可惜的是，对这些不同情况的精确方程并没有那么容易推得。

---

[18] 这个观察与在功率放大器这一章表述的下面这个观察十分相符：为大功率发射机的测试提供一个非辐射负载的简洁方法是用一根很长的同轴电缆。这个累加的损耗是足够地大以防止明显的反射，这样就保证了好的匹配。同时，这个损耗的分布特性隐指功耗是在一个大的体积内发生的，因而就减轻了发热的问题。这个窍门容许测试可以在没有专门设计的能承受大功率的电阻负载条件下进行。

[19] 又一次，由于辐射与任何其他损耗机理引起的功耗会导致在阻抗上的变化波动消失。

[20] 可以被证明绝对的理论最大辐射电阻的增强系数是 4，这对应于一个沿着偶极子的恒定电流情形。在实际上，所能取得的增强因子要大大低于被理论预测的值。

但是，对于 L 形天线这个特别的情形下(图 21.8)，可以通过下面的假设得到近似的公式(在这里需要风衣了，指假设过于大胆，译者注)：假设当天线被弯成 L 形时沿直导体的电流分布只是受到了相当小的影响。如果这一大胆的假设能够成立，那么可以得到相关的公式：

$$R_r \approx 40\pi^2 \left(\frac{l+d}{\lambda}\right)^2 \tag{15}$$

其中 l 和 d 如图所定义，而且假设总长度短于波长。这一方程是做了如此大的近似后得到的，以至于天线的长度需要被剪裁到一个合适的长度。尽管如此，这依然是一个用来建立初步设计的尺寸的较为合理的指导。当然，使用天线分析程序，比如说 EZNEC，减少了所需要的实际迭代次数。

如果水平部分的主要目的是增加电容，那么进一步改进可能需要增加更多部件。常见的(对 T 形天线来说)可以使用 2、3、4 根水平导体，这些导体相对于竖直部分呈对称分布。电容帽可以被认为是使用无穷根径向的导体棒的极限，见图 21.9。其他容性结构，比如球面或者扁球体，也被用来代替图中的平面圆盘。

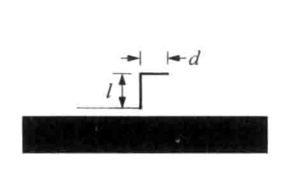
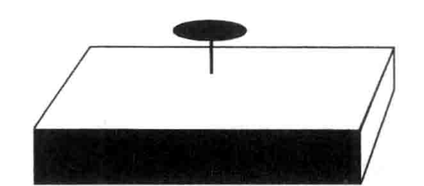


图 21.8　L 形天线


图 21.9　带有电容帽的天线

21.5.4　感性负载偶极子

只有当导体的尺寸与波长可比拟时，辐射才是显著的，在关于这点的讨论之后，下面这个结论可能会有些令人惊讶：偶极子所能得到的信号功率在任何一个单一频率下几乎和它的长度没有关系。这一不变性可以通过观察短偶极子传输的电压也低这一点来理解。在一阶近似下，可以通过计算天线长度和接收到的电场强度的乘积来计算开路电压，因此电压随短偶极子的长度而线性变化(直到某一个程度)。同时，我们已经看到辐射电阻随长度的平方变化。因此电压平方比上电阻是与长度无关的。随着偶极子长度减小，更小的戴维宁(Thévenin)电阻(辐射电阻)上得到越低的电压，使得可得的功率保持常数(忽略损耗)。比如，对于无损的单极子，可用功率为

$$P_{av} = \frac{(E_{pk}l)^2/8}{40\pi^2\ (l/\lambda)^2} = \frac{(E_{pk}\lambda)^2}{320\pi^2} \tag{16}$$

显然，可用功率与长度无关，而是取决于与场强和波长。因此，对于假设的无损偶极子，理论上说可以用无限短的导线来制作天线。这一结论看起来和我们的日常经验不一致，因为普通导线的辐射在低频电路分析中往往可以被忽略而不会造成问题，并且 AM 广播站也使用由非常高的塔来支撑的大尺寸天线。解决这一明显的似非而是的难题需要考虑到辐射电阻与驱动源的等效戴维宁电阻形成了一个电压分压器，而这个戴维宁电阻又被任何电路中都存在的电阻性损耗所增强。对于那些天线长度只占波长很小一部分的频率，辐射电阻相对于电路中的连线电阻相当小，以至于几乎没有多少功率能被传输到辐射电阻上。正是

由于这个严重的阻抗失配，常常使我们很方便地忽略了作为互连线的短导线辐射的可能性。任何导线在理论上任何时候都能够产生辐射，但是如果它的电长度很短那么这个阻抗失配就阻止了可以感觉到的辐射。这就是为什么导线知道什么时候是，什么时候不是通常电路中的天线。

然而，假设可以避免这种阻抗失配呢？毕竟，阻抗变换器是很容易被设计出来的。我们可以使得天线无限短么？回答是一个有限制的是。在认识到带宽和归一化的天线长度实际上是相互耦合的这一点之后，有一条限定条件就可以被认识到了。因为短偶极子有一容性的电抗分量，加入合适的额外感抗分量就可以使得天线电路在一个要求的工作频率下进入共振。电学上来说，只要电抗项消失了，天线就看起来变长了。这些负载电感通常被放置在偶极子的基座上（即信号馈入点），或者是在偶极子中心附近。但是，随着天线的减小，它的电容也减小了。为了保持共振，这个补偿的（外加的）电感必须增加。回想一下，一个串联谐振电路的 Q 为

$$Q = \sqrt{L/C}/R \tag{17}$$

很显然，Q 随着偶极子天线缩短而增加（假设无损），这是因为电感与电容都是大约与长度成正比的，并且辐射电阻与长度的平方成正比。因此，

$$Q = \sqrt{L/C}/R \propto l^{-2} \Rightarrow B \propto l^{2} \tag{18}$$

带宽因此与长度的平方成正比。结果是，可容许的天线的缩减是由所希望的通信带宽所限制的。进一步说，带宽越窄，天线的中心频率对周围的物体越敏感。如果要有意加入串联电阻来减小敏感度提高带宽，则往往带来了不希望的损耗的增加。即使没有故意加入电阻，损耗也总是存在的。如果要保持高效率，代表了这一损耗的额外串联电阻必须相对于辐射电阻较小。为了强调这一实际中包含的困难，考虑长度只为波长 1%的单极子天线。因此，辐射电阻大约在 0.04Ω。不用说，要让所有 RF 损耗比 40 mΩ 更小是极度困难的！因此在效率和带宽之间的基本折中是由偶极子能变短的程度严格限制的。带宽、归一化长度和效率之间的耦合使得大部分的天线设计至少有 10%波长的长度。实际偶极子天线几乎很少有小于这个值的，除非有效的信号功率十分大，足以容忍效率低的天线。

有时，会遇到有些天线同时使用容性和感性负载（即容性帽加上负载线圈）。这种额外引入的自由度可以让我们在做折中选择时有一定程度的放宽，但不能完全消除这种折中的约束。

21.5.5 磁场环形天线

电偶极子的对偶是由电流回路形成的磁偶极子。正如偶极子天线主要对电场很敏感一样，环形天线（图 21.10）主要对入射波的磁场分量很灵敏。我们马上就会看到在电偶极子天线面临很严重的问题的场合下，这一对偶性让环形天线在许多情况下很有吸引力。特别是在低频下，环形天线设计往往比它的对偶（电偶极子）更加实际，这就解释了比如为什么环形天线几乎为所有手提的 AM 收音机和许多寻呼机使用。



图 21.10 环路形天线（这里所示的是带有空气芯的版本）

下面关于环形天线有效辐射电阻的方程假设了其直径相对

于波长很小，而且没有使用磁性材料作为线圈的芯[21]：

$$R_{\mathrm{rad}} \approx 320\pi^4 \left(\frac{nA}{\lambda^2}\right)^2 \approx 31\,200 \left(\frac{nA}{\lambda^2}\right)^2 \tag{19}$$

其中 n、λ 和 A 分别是线圈匝数、波长和环路面积。

正如短电偶极子天线会产生一个净容性的电抗那样，磁环路天线具有一个净感性的电抗。Wheeler 的著名公式可以用来估计磁芯为空气的环路电感：

$$L \approx \frac{10\pi\mu_0 n^2 r^2}{9r + 10l} \tag{20}$$

这个式子假设了所有尺寸和电感抗为国际单位制，而不像 Wheeler 原始的公式里采用英寸为单位来得到微亨为单位的电感。

注意到相比电偶极子中这里多了一维自由度这一点十分重要：人们可以用更多匝数来增加辐射电阻。这一改进几乎和传统变压器中的阻抗变换完全一样。输入电波的变化磁场在每一匝引起相同的电压，因此在天线端得到 n 乘以每匝上感应的电压。由于能量必须守恒，电流就必须除以相同的因子 n，并且因此阻抗(这是电压与电流之比，欧姆教授如是说)增加了 n^2。因此即使环路面积比 λ^2 小得多，仍然可以将辐射阻抗增大到可用的值。

我们现在可以看环路天线是怎样解决 AM 收音机接收器这一棘手的问题的。AM 带宽最低频率处的信号其波长大约为 600m。任何手提的设备所允许的最大尺寸都必然只是这个波长的极小一部分。标准的与人体大小差不多的电偶极子天线都只能具有小得可怜的辐射电阻，在实际中不可能有效地工作。环路天线则提供了一个受欢迎的替代方法。主要是由于这个原因使得环路天线成为了可移动 AM 收音机中唯一使用的天线。进一步的改良是将线圈缠绕在一个铁氧体棒上来提供较大的磁导率以集中磁场。这种环路棒天线占领了可移动的应用，直到频率达到铁氧体的损耗大得使其不再可用(也许可以达到 VHF 的范围)。环路天线也是寻呼机的一个选择，因为这种终端设备要求天线具有非常小的形状因子(form factor)，这使得实现一个有效的电偶极子非常困难。而环路天线可以方便地设计成长方形，并安装进寻呼机的外壳里面。

21.6 微带贴片天线

已经看到只有当导体的尺度达到了与波长相当比例的大小时，辐射才成为可能。这并不总是我们想要的，比如因为辐射的损耗增加了微带线的衰减。尽管在这个应用下并不希望发生，这种辐射恰好是作为天线所要求的。当采用微带线制作时，这种辐射器被称为贴片天线，之所以这么命名是因为它们的形状。由于它们的平面特性而变得特别流行，这使它们与印刷电路一样在批量生产时十分廉价。尽管存在着不少严重的局限性(例如过大的 Q 值，即过窄的带宽)，它们的便携性和集成度在许多应用中补偿了其不足。

如图 21.11 所示，贴片是由一个矩形的导体片(在接地平面之上)来实现的。在一阶近似下，贴片天线被认为是将排在平面上的偶极子天线平行阵列并联成片的极限情况。因为

[21] A. Alford and A. G. Kandoian, “Ultrahigh-Frequency Loop Antenna”(超高频的环形天线), *Trans. AIEE*, v. 59, 1940, pp. 843-848；同时见 Kraus(前面已引到，脚注 3)。这个公式也假设导体的损耗为零。

这个原因，主要的辐射沿垂直于贴片表面方向，见图 21.12，其中贴片处于 x-y 平面上。

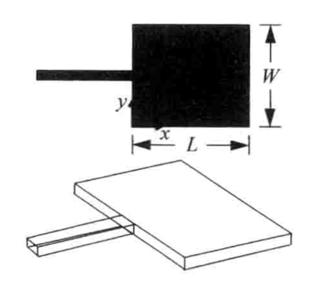


图 21.11　半波长贴片天线(导电层的形状与立体视图)

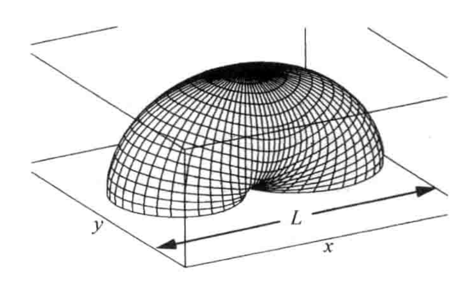


图 21.12　图 21.11 所示的贴片天线的典型辐射图形

辐射模式的确切特性可以通过控制如何来激励天线在很大范围得以调整。在大部分基本的配置中，贴片是在一端激励的，而且在某一边的中央(如图 21.11)。但是，也可以使用非中央的激励(偏移激励)来激发线性极化之外的模式。这一功能是非常有价值的，因为在许多微波通信系统中采用极化来提供缓解多路径干涉的方法[22]。也可以将激励线通过容性耦合到贴片，这可以在天线的横向边上或者是在贴片下加入额外导体层来实现。

对于图 21.11 所示的天线，假设长度 L 被取为半波长，那么电流就如经典的电偶极子天线一样在 $x=0$ 和 $x=L$ 处为 0，而在 $L/2$ 处达到最大。同时，电压在 $L/2$ 处达到最小，而在源端和最远端达到最大，这也与经典偶极子相同。

由程序 Mstrip40 仿真得到的电流密度如图 21.13 所示，该分布揭示了电流密度确实如我们所预测的那样以正弦变化，至少在肉眼观察下是如此。注意微带激励线的电流在沿线方向几乎保持着恒定的幅度，这表明已经达到了很好的匹配。这一图与图 21.11 中的贴片朝向是一致的。

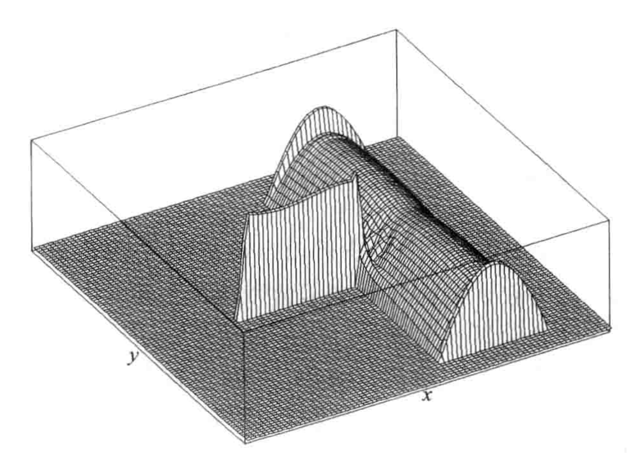


图 21.13　对于一个 1GHz 的矩形贴片天线，在中心频率下的电流密度分布(由 Mstrip 40 仿真得到)

---

[22] 从一个物体上的反射会改变极化的取向，比如说从逆时针极化转变成顺时针极化。这样，采用能有选择性地排斥这些分量中的一个就可以减小对带来麻烦的反射而造成的通信链接的影响。

贴片天线的重要特点是它相对较窄的带宽(典型值在 1% 的量级)。为了理解这一属性的来源,将贴片不仅仅想成电偶极子的平面阵列是很有用的。具体地来说,考虑结构是一个缺少六面中四面边墙的空腔振荡器(盒)。这么看的话,天线的辐射就成为了一个共振空腔中所泄漏能量的结果(具体地说,辐射主要是由于能量从宽度 W 的两个侧面中泄漏)。因为通常的介质层厚度相比贴片别的尺度来讲非常小,泄漏的能量相对于存储的能量要小得多。从而导致高 Q 值,使得贴片天线有很强的能力来过滤其他频率信号,但是它也要求设计和制造中更加精确,要求材料性质特别稳定。FR4 材料的变动性几乎肯定要求设计过程的多次试验,因此如果要用普通的贴片结构的话,它们被局限在对产品原型的制作或供业余爱好者使用。

因为要使得更多的能量泄漏出来以降低 Q 从而增加贴片天线的带宽,一个简单的方法是使用更厚的衬底材料。然而获得合适厚度的廉价低耗 PCB 板是十分困难的,这也有时激发了天线设计者使用空气作为电介质的主要部分(甚至全部)。百分之几十大的带宽都是因为有了这种设计才成为可能。在零阶近似下,对 VSWR 为 2:1 的分数带宽(忽略损耗)可以通过下面的经验公式很近似地得到[23]:

$$B \approx \frac{\varepsilon_r - 1}{2\varepsilon_r}\frac{WH}{L\lambda} \tag{21}$$

记住辐射主要在两条宽 W 的侧面上发生,我们可以预计能量泄漏,进而带宽,与相应的开口面积(WH)成正比,这就解释了为什么分子中出现这一因子。另一方面,增加 L 会增加存储的能量(空腔增加了尺寸),因此它像所示的那样出现在分母中。

为了强调这中间的困难,我们来看一个具体的数值例子。假设有一个 1GHz 的贴片采用 1.6mm 厚的无损电介质,其宽长比为 1.8∶1,相对电介质常数为 4.6。那么,估计的分数带宽达到大约 0.7%。这么小的值强调了采用贴片天线面临的挑战之一,因为生产中任何细小的误差都会导致天线不能在所希望的频率下有效率地辐射。

我们已经推迟了给出贴片天线辐射电阻的方程是有原因的,这是因为在文献中有着大量的设计方程,包括不同的复杂程度。遗憾的是,许多公式在一阶精度上彼此就存在着不同之处,因此读者要小心不要太过信任它们。为了在保证实用的前提下保持最大限度的直观理解,这里给出的方程是简单的(但是十分近似)。进一步说,已发表的公式大部分忽略了介质损失。可惜的是 FR4 很少是一个无损的材料,因此不得不修改公式来减小对那种材料的近似而引起的误差。

正如之前指出的,经典的贴片天线被设计为半波长辐射器,因此它的电长度被取为刚好等于半波长:

$$L_{eff} = \lambda/2 \tag{22}$$

在将电长度和实际长度联系起来时,重要的是要考虑到边缘场和有效介电常数两者:

$$L_{eff} \approx \sqrt{\varepsilon_{r,eff}}\,[L + 2(H/2)] = \sqrt{\varepsilon_{r,eff}}\,(L + H) \tag{23}$$

---

[23] 这个式子是对最初由 D. R. Jackson 与 N. G. Alexopulos 提出的公式的修正。这个原始的式子在"Simple approximate formulas for input resistance, bandwidth, and efficiency of a resonant rectangular patch"(有关谐振矩形贴片天线的输入电阻、带宽与效率的几个简单的近似公式),*IEEE Trans. Antenna and Propagation*, v. 3, March 1991, pp. 407-410。在这个原始式子中,在分母中没有 2 这个因子,而是在分子上有一个 3.77 的因子。在公式(21)中作的修正使得这个式子的结果更接近于电磁场解法器给出与实验得到的值。

其中 H 是电介质的厚度，而每条边长的近似长度修正为 $H/2$，这与 7.9 节得到的一样。还存在着更为复杂的长度修正量，但是 $H/2$ 这个值对于实际应用来说已经足够。

有效的电介质常数由下式给出：

$$\varepsilon_{r,eff} \approx 1 + 0.63(\varepsilon_r - 1) \cdot (W/H)^{0.1255}, \quad W/H > 0.6 \tag{24}$$

这个公式与用于普通的微带线的公式是一致的(公式(23)的边缘修正同样是如此)。下面的公式略微好记一点：

$$\varepsilon_{r,eff} \approx 1 + \frac{5}{8}(\varepsilon_r - 1) \cdot (W/H)^{1/8}, \quad W/H > 0.6 \tag{25}$$

由于典型的贴片天线的宽度 W 相比电介质厚度 H 要大得多，因此有效的介电常数往往接近材料的介电常数(比如只低 5%～10%)。因此，许多参考文献中给出的设计公式很少有区别这两种介电常数的。

我们需要类似地考虑边缘效应对有效宽度的修正：

$$W_{eff} \approx \sqrt{\varepsilon_{r,eff}}\,[W + 2(H/2)] = \sqrt{\varepsilon_{r,eff}}\,(W + H) \tag{26}$$

继续我们的设计方程，得到：

$$f_0 \approx \frac{1.5 \times 10^8}{L_{eff}} \tag{27}$$

在这个共振频率下，驱动点的阻抗理论值为[24]

$$Z_0 \approx \frac{90\left[\varepsilon_{r,eff}\dfrac{L + H}{W + H}\right]^2}{\varepsilon_{r,eff} - 1} \tag{28}$$

注意，宽长比对贴片天线的边缘看过去的阻抗有很大的影响。对于在有效相对介电常数为 4.2 的 1.6mm 厚的 FR4 上制作的贴片天线，这个公式表明 50Ω 的馈电点阻抗需要 W/L 值略微大于 3，而正方形贴片边缘阻抗大约为 500Ω。基于通用电磁场求解器 Mstrip40 的更严格的计算表明，W/L 值大约为 4.7 才能产生 50Ω 的边缘阻抗。而采用专门用于天线设计的程序(Pcaad 2.1)基于泄漏空腔模型还得到了更大的值，其比值超过 5。因此可以看到这些权威的软件也不能达到完全的一致[25]。

至少它们都认同一个基本的规律：宽度需要相当大才能达到 50Ω。希望宽度和长度相差不大以避免激起宽度方向的模式，这些模式会影响极化以及输入电阻。在实际设计中，宽度不应该被选择到大于长度的两倍，而通常它们应该被设成近似相等。

当考虑不可忽略的材料电介质比如 FR4 的损耗后，情况会变得不那么清晰。因为电介质损耗等价于一个并联电导，实际的贴片天线呈现的驱动阻抗比无损的模型中所预测的要小。为了考虑电介质损耗的效应，回想之前在有损电介质的电容中，在两端口之间的有效并联电导简单的为：

$$G = \omega\varepsilon(WL/h)\tan\delta \tag{29}$$

当极板尺寸远小于波长时，每块极板都能够被看作是等势面，这时这个电导是对电容器两端口进行测量得到的。贴片天线很明显不满足这个条件，因此必须加入一些额外的计算来得到正确的结果。在半波贴片天线上，电压基本上沿长度方向成余弦函数，在驱动端和最远处

---

[24] 如前所引的 Jackson 与 Alexopoulos。

[25] 话是这么说，重要的是要注意到 Mstrip 40 的预测是更接近于作者自己在实验室的经验。

边界处达到极值。同时忽略任何在宽度方向的电压波动。结果得到电介质中的功率损耗在 $L/2$ 处比两端要小。因此可以预计有效电导(从驱动源看过去)要小于 G。现在对于幅度为 V,随时间正弦变化的电压,可以知道电阻 R 上的功耗就是简单的 $V^2/2R$。类似地,在空间中以正(余)弦形式变化的电压产生的功耗也是 $V^2/2R$,其中 V 为驱动端的电压。

因为这个空间变化引入了在功耗上 2 倍的衰减(相对于电压为常数的情况),可以用一个连接在驱动端和地之间值为 $2/G$ 的并联电阻来表示总的电介质损耗。因此半波天线的总驱动点阻抗为

$$Z_{\text{ideal}} \,||\, \frac{2}{G} = Z_{\text{ideal}} \,||\, \frac{2}{\omega\varepsilon\,(WL/h)\tan\delta} = Z_{\text{ideal}} \,||\, \frac{2h}{\omega\varepsilon WL\tan\delta} \tag{30}$$

采用 FR4 得到的经验告诉我们,其典型损耗正切值为 0.022 导致了 W/L 值接近于 1.8 来在 1GHz 半波贴片中产生 50Ω 的馈电点电阻(W=12.6cm, L=7.0cm),意味着宽度必须要相对无损情况减小 2.6 倍来提供匹配[26]。除了这样产生了更可以被接受的占地因子(这是以效率的降低为代价的),这个损耗也使带宽由无损情况下的大约 0.7%增加到 2%(这里再一次相当任意地取 SWR<2 为带宽的定义)。正如之前所说的,窄带宽意味着——别的微带结构基本也是这样——需要反复的尝试来产生满意的贴片天线。

我们已经注意到采用大宽度的贴片来提供较低的驱动阻抗意味着要冒很难预测的天线行为的风险。因此贴片往往在很多应用中被做成方形的,这就必须要求有阻抗变换。一个经典的方法是在源与天线之间插入一段 $\lambda/4$ 的传输线。如果这段传输线的特征阻抗等于源与负载阻抗的几何平均值,那么就实现了匹配。

作为一个具体的例子,假设需要设计一个在 2.5GHz ISM(工业、科学与医学)频带中工作的供移动应用的 50W 的贴片天线。利用相当于 FR4 但无损的介质,矩形贴片尺寸大约为 27mm×80mm。这个长度非常合适,但是宽度与许多移动设备所要求的占地因子不相配。无论如何,3:1 的宽长比是很有可能激发出横向模式。假设选择了正方形贴片来替代,它的阻抗在理论上可以达到 500Ω。一根四分之一波长的 160Ω 的传输线可以实现相应的阻抗变换的要求。但这样大的特征阻抗必然会要求很窄的传输导线,从而制造工艺误差的控制就变得十分重要。这种尺寸的实际天线所呈现的阻抗更可能是上面值的一半那么大,这就缓解了对工艺控制的负担,此时所要求的匹配传输线阻抗接近于一个十分容易达到的 100Ω。注意到我们并没有关心阻抗变换的带宽,因为贴片本来就是频带十分有限的元件。

还有另一个阻抗变换方法可供贴片选择(实际上,对任何共振天线都适用)。因为在天线中存在的是驻波,电压电流都沿着贴片而变化。在我们所研究的半波天线的例子中,边界条件要求电流在驱动点和远端边界达到最小,而在中间达到最大。与此同时,电压在中央达到最小,在源与远端达到最大。阻抗作为电压与电流的比例,沿着天线而变化——从在驱动点的最大到中点处的最小——这与作为一个终端开路的半波传输线一样。通过使用嵌入驱动来利用这一阻抗变化的特点,如图 21.14 所示。

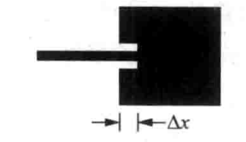


图 21.14　带有阻抗变换嵌入馈送的贴片天线(顶视图)

---

[26] 又一次,模拟程序 Mstrip 40 的计算结果与这个经验十分相符。

作为一阶近似，我们再一次将结构看作是传输线。因此还是假设驻波（电压和电流）沿着贴片的长度方向正（余）弦变化（可以回想起来，这些假设和在推导普通偶极子天线的辐射阻抗时完全一致）。电流在贴片的两端近似为零，向中央大致按正弦变化。同时，电压在两端为最大，向中央以正弦波形式衰减为零。因此，当驱动点向中央移动时，电压电流比近似二次方的变化，因为电压正弦衰减同时电流按同样的速度增加。因此阻抗可以通过乘以下面的因子得到：

$$Z \approx (Z_{\mathrm{edge}})\left(\cos\pi\frac{\Delta x}{L}\right)^2 \tag{31}$$

其中 Z_{edge}是没有嵌入驱动时的驱动点阻抗 s。

对于这个难以把握的无损 FR4 的贴片天线的例子来说，我们需要向下变换一个相当大的因子，大约为 10.8，即

$$\Delta x \approx \frac{L}{\pi}\arccos\left(\frac{1}{\sqrt{10.8}}\right) \approx 0.4L \tag{32}$$

注意计算得到嵌入位置为几乎要凹进到中心。因为在接近中心处，电压随距离的变化大[27]，在这个嵌入位置附近的阻抗的精确值对于距离非常敏感。加上我们的近似中内在的不确定因素，意味着获得正确的阻抗值需要大量的反复调整。另一个需要考虑的问题是如此深的嵌入代表了一个一阶微扰，我们应当预计到极化和其他感兴趣的参数的变坏。对于实际的 FR4 来说，电介质损耗会降低正方形贴片的边缘阻抗，意味着一个一般大小的阻抗变换比就足够了。

一种方便的为实现原型的裁剪嵌入式贴片的方法是先形成一个比计算结果更深的嵌入。然后通过放置一个短的条来覆盖嵌入空的某些部分来简单地上调阻抗，见图 21.15。这一特别的方法避免了精确切割的要求，同时便于多次的迭代。焊接（与去焊接）一片铜箔带要比在铜覆盖层上刻掉一部分要来得简单得多。

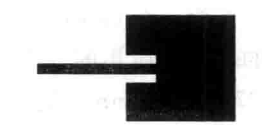


图 21.15　对嵌入式贴片天线的调整方法（顶视图）

内嵌式驱动的一个缺点是它会扰动场的分布一个量，这个量随着凹口的深度而增加[28]。三种阻抗匹配的方法（控制 W/L，$\lambda/4$ 变换器，嵌入式驱动）提供了在感兴趣的参数间进行折中的重要的自由度。比如说，如果矩形贴片的尺寸产生了 200Ω 的阻抗，就可以采用内嵌式驱动把阻抗降低到 100Ω，然后通过 71Ω 的 $\lambda/4$ 传输线达到 50Ω 匹配。由于每一步的阻抗变换都只包含相对较小的比率，因此可以得到更为实际和鲁棒的设计[29]。再一次指出，FR4 的损耗从某种程度上降低了对急剧阻抗变换的需求（同时并展宽了带宽，虽然这是以效率降低为代价的），但是基本的贴片天线仍然是一个窄带结构。

另一种替代嵌入式驱动的方法是保持贴片本身不受改变，而通过在合适的接点上由背

---

[27] 电压在中心处是达到最大，但是其空间的导数在那里是最小。

[28] 这个敏感度却是常常被利用的。一个恰当的定位的偏离中心的馈送会同时激励横向与纵向的模式，而它们的叠加产生圆周状的极化。

[29] 当然，我们也能够用一系列的 $\lambda/4$ 变换器或某些其他的变种。存在着许多实现所需要的变换的方法，读者可以独立地寻求这些替代的办法。

面的一个连接器来驱动。这个装置在图 21.16 中给出。尽管这样的方法要求有一个通孔（或者，更常见的是适合穿过短导线的小孔），这个方法并不算麻烦，因此大部分商用贴片天线都采用了这种设计。但是，必须要意识到同轴线驱动的两个潜在问题：如果采用相对较厚的衬底来增加带宽，那么驱动线的电感可能会是很大的，从而改变了调谐和阻抗。同时，驱动线可能自身会产生一定的辐射，导致整体辐射模式的畸变。

一种显然的解决因驱动线电感引起的匹配问题的方法是采用传统的阻抗匹配网络。但是，另一个有趣的替代方法是调整贴片本身的几何尺寸来提供合适的补偿容抗。比如说，假设在驱动点周围开槽以使得在驱动点附近的区域产生一个近似的平板电容，见图 21.17。

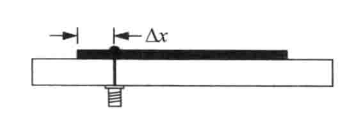


图 21.16　带有同轴馈送的贴片天线（侧面图）

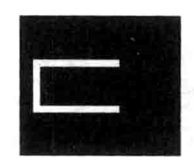


图 21.17　带有 U 形槽的贴片天线（顶视图）

当 U 形槽在合适的位置和尺寸被刻出时，某个出乎意料的事情发生了：2∶1VSWR 带宽陡然增加了原来的 40%[30]。这一结果与墨菲定律通常的预测（结果总是比我们想象的要坏）完全相反，这是因为产生了多个共振模式。不仅仅存在着基本的阶贴片共振模式（我们可以称其为共模共振），在驱动感抗与 U 形电容之间还存在共振。同时也存在一种差模共振，在这种模式下与 U 形边界两边相对的贴片部分反相的工作。如果这些不同模式的间隔有合适的量（比如像在 Chebyshev 滤波器中那样），那么总的带宽相比基本的贴片可以大大增加。

从上面这些定性的描述，你可能会得到一个印象即设计来达到这个神奇的条件是相当困难的，你的猜想的确是对的。目前还不存在简单的解析公式来设计这些天线，在设计汇聚可以接受的结构前需要用场求解器进行许多次的迭代。就像工程中所有其他的东西一样，在这个基本的设计要素（比如驱动的选择、边界的形状）上有许多变种，这样就进一步扩展了搜索的空间。这种可能的行为的多样化研究激起了在这一领域的活跃的研究。比如，除了在宽频带下相对不变的增益，有时希望在几个分别对应分离通信带的分立频率附近存在峰值。尽管没有突出的开槽形状规范，这一原理已成为了贴片天线广泛应用于许多多频带手机的基础，尽管非明显的（非 U 形）的槽形状依然是个通常的选择。

宽带贴片天线

U 形槽的修正使得贴片天线的带宽增加了一个数量级。尽管十分令人吃惊，但是许多应用中即使 30%～40%（相对于载波频率而言）的带宽仍然是不够的。比如，最近对超宽带（UWB）通信的兴趣就要求天线能够在 3∶1（甚至超过 10∶1）的频率范围工作。遗憾的是要获得这么大带宽的天线需要探索超出对标准天线简单修正的天线的概念。

在讨论某些超宽带天线之前，来搞清楚几个概念是有帮助的。带宽这个词是有一点歧

[30] T. Huynh and K. F. Lee, "Single-layer single-patch wideband microstrip antenna"（单层、单贴片宽带微带线天线），*Electronics Letters*, v. 31, no. 16, 1995, pp. 1310-1312.

义的，比如说，因为它可以指阻抗带宽，或者是定向性保持在某一特定值之上的带宽，或者极化值限制在某些误差以内的带宽，或者是主辐射叶瓣在某一最小宽度以下的带宽。基于其实际内容，一个或多个定义可能会有关联。理想的 UWB 天线在整个带宽范围内存在一个不变的辐射图形、增益和阻抗。同时，有些 UWB 系统（比如脉冲调制系统）额外要求在同样宽的带宽上的色散较低。至少这么说，同时达到所有的目标显然是很有挑战性的。

在设计 UWB 天线时的一个有用的指导原则是有关一个具有特定导体图形的天线阻抗和与该图形互补的另一个天线阻抗的关系。通过对这些互补结构应用 Babinet 原理[31]，我们发现这两个阻抗通过自由空间的阻抗而互相关联：

$$Z_1\overline{Z}_1 = \eta^2/4 \tag{33}$$

其中有上划线那个符号代表有互补导体图形的阻抗[32]。

现在假设能够设计出一种天线，它的导体图形是其本身的互补，那么 $Z_1=\overline{Z}_1$。在我们可以宣称成功了的程度内，预计天线的阻抗（在自由空间里）大约为

$$Z = \sqrt{Z_1\overline{Z}_1} = \eta/2 = 188.5\Omega \tag{34}$$

也就是说，这个天线在理论上在无限宽的带宽里有一个常数的阻抗。现在，要满足这一条件在实际上是不可能的（因为如果你想一想，照字面意思，必须要求导体无限大），但是寻求它的近似就导致了实际有用的天线。如果天线是被制造在电介质材料里或之上，那么自由空间的阻抗值就会被缩小一个有效的相对介电常数的平方倍。即使考虑了这个因素，许多 UWB 天线仍然表现出低于预计的这个“神奇”阻抗（这主要是因为实际实现并不是完全自我互补的）。典型的情况是，可以期待实际天线的阻抗值是基于完美互补假设计算得到的理论值的 60%～80%。

我们对 Babinet 原理的关注并不表明自我互补的天线是唯一的超宽带天线。只是因为这一原理对超宽带的应用使得我们能够以一个概念上简单的方法来汇聚设计是努力以产生许多种的宽带天线。一个具体的例子就是经典的基于指数关系的等角度（或对数）螺旋图形，这是基于以下的指数关系：

$$r = r_0 e^{\phi/\phi_0} \tag{35}$$

这个式子也可以表示为

$$\phi = \phi_0 \ln(r/r_0) \tag{36}$$

其中 r_0 是角度 ϕ 为零时的半径，而 ϕ_0 则控制了半径随相位角增加的速度[33]。ϕ_0 的典型值在 4～5rad，也就是说每转一圈半径增加大约 4 倍。这个因子相应也就被称为扩张比，而 ϕ_0 有时也称为外展角（它的倒数为外展率）。

之前的这些方程定义了一般的导体形状，但并没有完全定义导体的图形。螺旋图形的每一支有两个边界，它们的一般形式都由公式(35)刚才，但互相间偏离某个角度。更进一步

---

[31] 由 19 世纪的法国物理学家 Jacques Babinet 提出与阐明。

[32] Babinet 原理对你来说可能从物理课的有关光学的部分更为熟悉，这是作为一个简化某些经典的衍射分析的方法，但是这个原理被应用到天线的设计问题无疑是更为重要。简单地说，假若我们由 ηH 来产生 E' 或是从 E/η 产生 H'，得到的经变换的场（标为有撇的变量）依然满足麦克斯韦尔方程，这是有无后者具有的对称性质。作为一个结果，将一个导体图形用其互补结构取代会产生我们得到的阻抗关系。

[33] J. D. Dyson, “The equiangular spiral antenna,”（等角度的螺旋天线）*IRE Trans. Antenna and Propagation*, v. 7, October 1959, pp. 329-334。

说，实际的考虑排除了将各支扩展到半径为无穷，因此必须颇为任意地决定如何与什么时候来停止扩张。通常来说，最好的结果（以在通带中阻抗为常数来度量）是通过将末端逐渐变细来得到，但是截断则是比较容易实现的（如图 21.18 所示）。

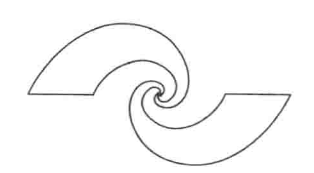


图 21.18　双臂、1.5 圈截断的等角度（对数）螺旋天线

最大的半径决定了低端的频率极限。作为一个十分粗略的近似，低端频率波长的极限大约为最大半径的 4 倍左右。高频极限是由螺旋体驱动点的半径 r_0 决定的，频率上限的波长极限大约等于这个最小半径的 4 倍[34]。

从实验的角度来看，1.5 匝螺旋天线对工艺误差不十分敏感。因此，图 21.18 的版图就用了这个匝数。这个主意用 FR4 实现的一个特定的例子给出了回波损耗当频率在 400MHz 到 3.8GHz 时优于 10dB，同时在这个频带内具有 80Ω 的阻抗[35]。这一十倍频的范围，尽管十分出色，但并不意味着绝对的极限。从设计原理来看，我们能知道可以（通过增加螺旋的半径）降低下限，（通过减小最小半径）提高上限。为了保证全部带宽潜力的实现，必须在连接驱动端时十分仔细。电抗不连续和其他寄生效应会很容易将带宽降低到理论最大值的很小的一部分。

另一种自我互补的天线并不基于指数而是线性的螺旋体。在这些阿基米德螺旋体中，每个臂的边界都是基于具有以下基本形式的公式[36]：

$$r = r_0 \phi \tag{37}$$

见图 21.19。对于指数螺旋天线来说，设计自由度包括天线臂数目、最小和最大半径、外展率。同样的通用设计标准可以采用。低（高）端频率的极限是由总的（最小）半径设定的，这两个半径都为 $\lambda/4$ 条件决定。

而另一种方法可以不用这些弯曲的几何形状而采用更为简单的曼哈顿版图。一种可能的结果是方形或者矩形的螺旋图形，如图 21.20 所示。就如读者会赞赏的，画一个方形的螺旋形版图要比产生一般的螺旋版图容易得多。这里的自由度包括天线臂的数目以及最大与最小半径。也可以选择一个线性或者指数的外展。最后，虽然方形要方便很多，其他形状显然也可以选择。



图 21.19　双分支 1.75 圈阿基米德螺旋导体图形的例子

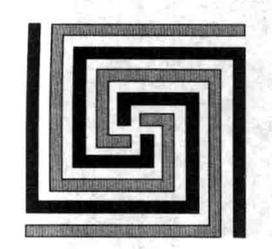


图 21.20　四臂正方形螺旋导体图形的例子

---

[34] Stutzman and Thiele，脚注③。

[35] J. Thaysen *et al.*，"A logarithmic spiral antenna for 0.4 to 3.8GHz"（适用于 0.4～3.8GHz 的对数螺旋天线），*Applied Microwave and Wireless*，February 2001，pp. 32-45。

[36] Stutzman and Thiele，见脚注③。

除了自我互补外，自我相似性也是在宽带天线设计中十分有用的指导原则。定性来说，认为在不同尺寸下的自我相似性会让天线有宽带的特性是合理的，因为这样做，在不同频率下天线会有相同的归一化电学长度。由于分形(fractal)图案是具有这种精确的自我相似性的几何体，分形图案天线在近几年吸引了相当多的注意力。但在这个上下文中，可以看到经典的指数周期(log-periodic)偶极子阵列就是一种自我相似的结构，因此分形图案天线也并不是像我们所想的那么新颖。也就是说，天线类型比其分类来得要早[37]。

21.7 其他各种平面天线

并不是所有的天线都能很好地归属到我们所给出的类型中去。经常，如封装、成本、频率范围这些限制使得书本里描述的天线无法使用，迫使进行大量的改动。比如，手机和其他消费产品的设计必须常常向美学因素让步。近来更常见的是，这些设备的天线必须根据封装而改变，而不是倒过来(根据天线来修改封装)。

一种宽带天线是基于19世纪Oliver Lodge爵士发明的双锥型天线来适应当时使用的由电火花产生的宽带谱的信号(见第1章Who Really Invented Radio)。用平面结构的形式实现，这个双锥形天线就变成了流行了几十年的领结型天线，它用于电视天线来覆盖一个两倍频率的UHF频带。用微带线实现的这种天线工作得也很好，同时因为实现起来十分简单而广为流行，见图21.21。

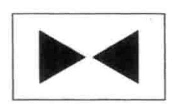


图21.21 用微带线实现的领结形贴片天线(顶视图)

这些三角形的贴片通过其顶点来驱动——或者是通过一对在表面上的微带线或者是通过处于不同导线层的导线——这就像任何别的贴片型结构一样。某些实验性的UWB系统采用领结型天线，在那里被称为“双鳍天线”，这或许是要给“领结”这个词所缺乏的技术气质增添一些风采。

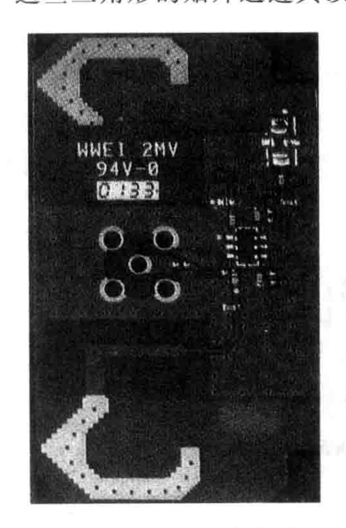


图21.22 802.11b无线局域网卡产品中所用的具有代表性的平面天线

我们应该提及所给出的任何天线结构也可以用作阵列中的单元。取决于个别的天线是如何设计和驱动的，可以进行波束成形或者宽带化(或者两者兼顾)。

作为另一个例子，考虑一个典型的无线局域网卡。第一代这类产品通常采用短电偶极子天线从卡平面垂直地突出出来。这类卡的性能非常出色，但是随着产品类型的进步，工程师通过转而使用天线平面来满足更进一步小的占地面积的要求。图21.22就是一例。

可以清楚看到，两个勾形天线并不能容易地被分类为贴片天线或者偶极子天线，尽管它们看起来更接近于后者。实际上，它们看上去不像是设计得很细致的。这

[37] 对于许多这些经典天线的讨论，见J. D. Kraus以及*The ARRL Antenna Handbook*(这两个参考材料都在脚注3中提到)。

种卡的生产厂商实际上是通过提供外加天线的接口来暗暗地承认了这一事实。在照片中可以看到，PC 板中存在着放置 SMA 连接器的钻孔，这样如果需要的话可以外接一个性能好的天线。作者拆卸了许多无线局域网卡，发现它们中大部分有这一暗藏的结构。额外加入 SMA（加上一到两个表面贴装组件）通常就可以容许外部天线的使用，而用外接天线可以使这些卡的有用通信范围增加 50%～100%。

21.8　总结

我们已经看到辐射本质上是光的有限传播速度的结果。为了产生合适的高的辐射电阻要求有足够的相移，这就解释了为什么实际的天线长度至少要达到波长的相当一部分大小。

基本的电偶极子天线（对称的或者单端对地的）不仅仅十分常用，它们也用作理解更复杂天线的重要基础。短的电偶极子有着较低的辐射阻抗并且主要是呈容性的。电容帽可以用来增加辐射阻抗，而电感则可以用来谐振掉任何电容。这些方法直到某一个极限都是有效的，而这个极限是为提供一个给定的最小带宽或最低效率的需要，以及产生一个其特性对几何尺寸或环境条件的小的变化不那么敏感的要求而形成的。这样做的折中得到的结果是长度远小于波长的十分之一的天线，其性能往往是不能令人满意的。

磁环路天线可以看作是电偶极子天线的对仗。不同于电偶极子天线，磁环路天线的辐射电阻与匝数有关，这对天线的紧缩尺寸实现带来了一个额外的自由度。

贴片天线可以看作是无穷多个无限细的偶极子天线的并联组合。但是，它过高的 Q 值在许多情况下是一个很大的缺陷。增加一个 U 形槽有可能增加一个数量级的带宽，使得它的设计可以允许更多的工艺误差。

最后，我们研究了一些超宽带天线，它们的设计是由受到 Babinet 原理启发得到的自我互补性来指导的。超过十倍频带宽的实际天线从而变为可能。

第22章 集总滤波器

22.1 引言

滤波器设计是一个如此广博的话题，我们并不指望能够在教科书的一个部分中把这个题目完全解释清楚。的确，即便选择把内容分成两章来写，这里的有限目标也不是尝试对所有可能的滤波器类型进行完整的回顾或是来试图提供其基础理论中的详细数学推导过程，而是集中在与滤波器相关的定性的主意和实用的信息上。对于那些对我们所忽略掉的严格推导感兴趣的读者，我们当然会提供寻找相关文献的链接。而对那些甚至希望忽略掉我们所提供的已经是十分有限而不严格的推导的读者，他们可以直接跳到本章的最后，在那里以“菜谱”的形式对部分常见的滤波器类型进行了总结。

尽管对平面工艺的强调通常意味着我们的讨论会局限于微带线的实现上，很多这样的滤波器是直接从低频集总原型上推论得来的。因为通过学习这些集总元件的原型，可以帮助掌握许多重要的概念，因此我们将大致按照历史的顺序，从集总滤波器设计开始讨论。不可否认，尽管集总式和分布式的模型在实现的细节上是有区别的，但是对于问题一些基本理解是相通的，这就是在本章中希望着重介绍的内容。

这里只考虑无源滤波器，这部分是为了把章节的长度限制在可控的范围内。另外一个原因是，与无源滤波器相比，有源滤波器通常有更高的噪声与非线性、有限的工作频率范围、更高的功耗和对参数变化相对来说更高的敏感度，尤其是对我们所关心的 GHz 的频率范围而言。

22.2 背景——一个简短的历史回顾

选频电路的应用至少可以追溯到对电磁波的最早期的研究。在 Hertz 的经典实验中，他正是使用了偶极子和环形天线（即环形振荡器）来净化由他的火花隙装置所产生的频谱，进而在他的十分原始的接收机中引入了一定程度的频率选择性。无线通信研究的先驱者 Oliver Lodge 爵士创造了“共振（Syntony）”这个词来描述调谐电路的作用，表现出对这种调谐特性的价值的有意识的好评，这个认识是在考虑到当时使用的火花塞信号具有的难以避免的宽带本质情况下得到的①。几乎与此同时，Nikola Tesla 和 Guglielmo Marconi 开发了他们自己的调谐电路（Marconi 的英国专利 #7777 价值的重要性是如此之大，这甚至导致了一场激烈而漫长的诉讼案②）专门用于排除不希望接受的信号，从而预见到基于正弦波载波

---

① 见 Hugh Aitken 的出色著作《谐振与火花塞（Syntony and Spark）》（普林斯顿大学出版社，普林斯顿，新泽西州，1985）。该书对无线通信的早期工作作了技术上十分详尽与精彩的回顾。

② 就如已经在第 1 章中提到的，美国高等法院引用了 Lodge，Tesla 与 Stone 的早期工作来拒绝 Marconi 在 1943 年的“4 个 7”的专利的美国版本。

的通信的到来。

然而，尽管有了这个基础，现代的滤波器理论并不是直接追溯到在无线通信研究中的那些早期努力。事实上，这方面研究的根基甚至要回溯到更久之前：正是对应用于电报（然后是电话）的传输线性质的研究构成了早期的滤波器理论。在 1854 年，William Thomson（他其后成为 Kelvin 爵士）对传输线首次进行了分析，当时是只考虑了传输线的分布电阻和电容。他作为对 4000km 的跨越大西洋电缆项目的顾问的工作，建立了实际传输率和传输线参数之间的关系③。二十多年之后，Oliver Heaviside 以及另外一些学者通过包括分布电感对 Kelvin 的分析进行了补充、增强，从而大大扩展了能够准确描述传输线特性的频率范围④。根据 Heaviside 工作的一个特定的推论，大约在 1900 年，美国贝尔公司（American Bell Company）的 George Ashley Campbell 与哥伦比亚大学（Columbia University）的 Michael Idvorsky Pupin 建议对电话传输线以一定的空间间隔加入集总电感来减小色散（即脉冲形状变模糊）⑤。这个建议是与滤波器理论相关的，因为 Heaviside 意识到集总线路区别于连续线路在于前者存在一个确定的截止频率。Campbell 和 Pupin 对于保证一定的最小带宽提供了设计指导⑥。

以工程可以实现的形式，集总式传输线的明显的有限带宽这个缺点很快变成了一个长处，进而建立了滤波器设计的一个主要的进化分支。第一个发表的公式是由 Campbell 提出的，1922 年他发表的经典论文更为详尽地讨论了他在第一次世界大战期间开发并申请了专利的想法⑦。与此同时，Karl Willy Wagner 也大约在同一时间发展了这些想法，但是德国的军事政府推迟了其发表，于是让 Campbell 占了先⑧。现在，大家承认这两位先驱者应当为独立并且几乎同时想到这个同样的了不起想法而共享这份荣誉。

Campbell 的同事，Otto J. Zobel 发表了一份经常被引用的对 Campbell 工作的扩展，但这依然是基于传输线的一些主意⑨。在之后几十年的发展时间里，我们可以看到一个对下述这些频率响应的不断深入的理解：即如何能够在实践中实现逼近理论上想见的理想的平坦通带、常数的群延时以及到无穷衰减的阻带的无穷陡过渡。当意识到这些理论上的理想

---

③ 他因为这个对这样一个伟大的工程的成功所作的贡献而被授予爵位，这个工程将“老”与“新”的世界第一次连接起来。只有一个船，即 The Great Eastern（大东方），是足够大到能装载及铺设重达 8 百万千克的电缆。经历了自 1857 年开始的四次令人沮丧及昂贵费用的失败，在 1866 年 7 月 27 日，终于用一根电缆将爱尔兰的 Valentia 与（加拿大）纽芬兰岛连接起来。这根电缆一直不断地承担传送电报的任务，直到 1965 年。

④ 要知道关于这个故事的过多的背景，参看 Paul J. Nahin 的极佳的著作：Oliver Heaviside: Sage in Solitude（Oliver Heaviside：一个孤独中的圣人），IEEE Press，New York，1987。

⑤ 就像许多具有重要商业价值的关键性主意，一场有关这一想法的争斗发生了。这是有关贝尔实验室已经在用由 Campbell 发展的远早于 Pupin 的 1900 年文章的发表的线圈进行实验了。Nahin（如前所述）观察到 Pupin 的自我推荐的能力实在是好，以至于他超前别人获得了专利。他甚至因为他这个“发明”从 Campbell 的雇主那里赚取了超过四十万美元的专利费（那时候美国还没有个人所得税）。更过分的是，Pupin 在 1924 年发表的获得普利兹（Pulitzer）奖的自传没有提及 Campbell 与 Heaviside 的贡献。

⑥ A. T. Starr, Electric Circuits and Wave Filters（电路与波的滤波器），2nd ed., Pitman & Sons, London, 1948。

⑦ G. A. Campbell, “Physical theory of the electric wave-filter”（电波滤波器的物理理论），*Bell System Tech. J.* v. 1, no. 2, November 1922, pp. 1-32。同时见他的美国专利 #1,227,113，1917 年 5 月 22 日提出。

⑧ “Spulen- und Kondensatorleitungen” [Inductor and Capacitor Lines]（电感器与电容器传输线），*Archiv für Electrotechnik*, v. 8, July 1919.

⑨ O. J. Zobel, “Theory and Design of Uniform and Composite Electric Wave-Filters”（均匀与综合的电波滤波器的理论与设计），*Bell System Tech. J.* v. 2, no. 1, January 1923, pp. 1-46.

情况是无法达到的，人们开始了解近似的必要性。其中一些更重要、更实际、更完善的近似包括我们在本章中会学习到的 Butterworth（巴特沃兹），Chebyshev（切比雪夫）和 Cauer（椭圆型）滤波器类型。到了 20 世纪 30 年代，Sidney Darlighton 等学者开发了直接综合滤波器的以达到特定的通带、阻带指标的方法。

在二次世界大战和之后的一段时间，滤波器设计这个领域以更快的速度发展。对于在集总滤波器的设计中容许有 Q 值为有限值的元件方法的研究提供了更好的可预测性和准确度希望。在微波的领域，直接基于集总元件的原型拓扑结构为那些利用（而不是忽略掉），分布效应的拓扑结构所补充。它们当中，很多被以微带线形式容易地实现。在下一章中会以此题目来进行重点讨论。

晶体管的兴起使得有源元件的尺寸不再是决定电路尺寸的主要因素。众多的有源滤波器拓扑结构出现微型化的要求，在许多情形下，取代体积大的电感-电容电路。除了能够有效地大幅度地减小尺寸，一些有源滤波器也是电可调谐的。然而，这些特性也需要付出代价：有源滤波器消耗能量、受到非线性及噪声的影响和为保证增益单元正常特性而导致工作频率上限的减小，这是因为需要在高的频率下实现具有良好控制特性的有源部件。这些折中考虑随着微波频率的接近变得更为关键。这并不意味着微波有源滤波器就不能在某些特定的应用中很好地工作（因为有许多成功的例子存在），但是在这种频率下最好的性能依然是用无源元件来实现的。出于这个原因，我们仅仅考虑无源滤波器。

晶体管的出现也与计算成本的迅速降低同时发生（并且起到了推动作用）[10]。不再受限只是考虑那些直截了当的解析解，搞理论的人可以从更一般的角度来提出滤波器近似的问题，例如，“在一个给定的频率区间，设置一个网络的极点与零点来使均方误差最小化（或误差最大，或某种其他的性能指标）以实现次品率在 1%之内，这是相对于一个理想的频率响应模板而言的，同时要考虑到元件误差的局。”采用经典的解析方法要兼顾这么多的因素是十分困难甚至是不可能实现的。然而使用这些现代的综合方法，得到的滤波器的性能对于给定的滤波器阶数在一定意义上是不可能做得再好了（这是就原始给定的设计标准来衡量的）。但带来的缺陷则是设计现在不能再用手工计算来完成，因而变得不容易去理解。我们这里的主要目标因此是为这些滤波器是如何工作的提供一个直观的解释，而将具体的执行过程留给机器计算来完成。由于这些直观的解释缺少细节，我们将要提供的公式主要是提供一个起点，依此出发通过迭代能够达到一个满意的最终设计。

22.3 用传输线构成的滤波器

我们从 Campbell、Wagner 和 Zobel 的“电学波滤波器”开始。如已提及的，这些是从对传输线的集总近似推得的，因此我们从讨论这种“人造的”传输线来看一看有限的带宽是如何产生的。

---

[10] 遗憾的是，由于篇幅的限制，不得不略去有关 Teledeltos 文章、电解质槽，以及其他一些基于势理论来设计滤波器的模拟计算机的有趣的故事。

常数 k 型("镜像参数")滤波器

为了方便,将第 2 章的一些计算在此重复。回想一下,首先考虑如图 22.1 所示的一个无穷长的梯状网络的输入点阻抗,Z_{in}。求解得到 Z_{in}

$$Z_{in} = \frac{Z \pm \sqrt{Z^2 + 4(Z/Y)}}{2} = \frac{Z}{2}\left[1 \pm \sqrt{1 + \frac{4}{ZY}}\right] \tag{1}$$

其中,无源性的考虑使得我们只选择实部为非负的解。

图 22.1　用作人造传输线的无穷长梯形网络

作为一点特别(但是典型)的例子,考虑一个 $Y=\mathrm{j}\omega C$,$Z=\mathrm{j}\omega L$ 的低通滤波器。那么,这个无限长的人造传输线的输入阻抗可以表示为

$$Z_{in} = \frac{\mathrm{j}\omega L}{2}\left[1 - \sqrt{1 - \frac{4}{\omega^2 LC}}\right] \tag{2}$$

在很低的频率,根号内的值为负且数值上较大,使得括号内的项几乎是个纯虚数。在这样的频率范围中,总的 Z_{in}因此基本上是实数,为

$$Z_{in} \approx \sqrt{Z/Y} = \sqrt{L/C} = k \tag{3}$$

由于$\sqrt{Z/Y}$为一个常数,这类滤波器通常被称作恒 k 滤波器[11],文献中这也被称为镜像参数滤波器。(对于如我们这里考虑的这类由重复单元构成的滤波器,存在一个叫做镜像阻抗的量,它与特征阻抗 k 相同。)

只要输入阻抗存在实部,非零的平均功率就会由源端耦合进传输线。然而,当高于某些特定频率,输入阻抗变成纯虚数,就如我们从公式(2)中可以看到的那样。在这种条件下,没有实际的功率能被传输入网络中,这样滤波器的衰减很严重[12]。因为这些显而易见的原因,对输入阻抗变成纯虚数的频率被称为截止频率,对于这个低通滤波器例子而言,截止频率由下式给出

$$\omega_h = 2/\sqrt{LC} \tag{4}$$

当然,任何实际滤波器必然采用有限个单元,因此很自然地会产生对那些假设无限重复单元结构滤波器的分析有多大价值的疑问。直觉告诉我们,只要有"足够多"数目的单元就能给我们带来能够接受的准确度。基于集总参数网络的理论,我们同样期望用滤波器阶数来控制最终的幅度滚降(rolloff)速率。因此所要求的滚降特性决定了网络采用的单元(或段)数目(稍后我们将就此有更多讨论)。单元的数目更多,滚降率越高。如我们将要看到的,也存在着某些(但是在实践上是受限的)Z 与 Y 的灵活度,允许在通带、过渡带和阻带特性之间作某种程度的折中。然而,基于人造传输线概念的滤波器仍然具有一个限制,那就是要在设计过程中直接结合对特性加的具体要求是难以实现的。注意,例如,通常忽略了关于滤波器在截止频率附近行为的讨论。我们并不清楚由通带到阻带的过渡是渐变还是突变

[11] Campbell 对 k 的用法仅限于这个上下文,但是 Zobel(见脚注 9),他很明显是第一个在发表的文献中用了恒 k 这个术语。

[12] 在不存在着带有能耗的元件情况下发生衰减,初看起来与直觉不相符。然而,考虑到一个滤波器可以通过反射能量而不是消耗能量的方式工作,即,滤波器可以处于有目的地通过产生在某个频率范围中阻抗失配的工作方式。现代的滤波器综合方法是直接基于调控作为频率函数的反射系数。

的，是单调的还是振荡的。我们同样不知道通带的精确形状。最后一旦过渡区的形状不符合要求时，我们没有如何去修正它的指导方针。就如之后将会看到的，这些缺点导致我们去考虑其他的滤波器设计方法。

一旦滤波器的阶数被确定了(不论是用什么方法)，下面一个问题就是如何来进行匹配终断。注意那个无限长的人造传输线分析假设了滤波器是被一个具有如公式(2)所描述的行为的阻抗终断的。就是说，假定存在的有限长的滤波器必须要被一个由原来的无限长的梯形网络所产生的阻抗终断，即在低频时具有一个实阻抗，而超过截止频率之后变为纯虚数。换句话说，要严格满足公式(2)给出的条件需要荒唐地以一个我们所希望得到的滤波器作为匹配。因此我们不会太吃惊来发现任何一个实际的实现都需要一定的折中，所有这些折中方法均与用一个有限长的结构来近似无限结构所具有的阻抗，这是一个不可能实现的任务。例如，一个很常见的选择是简单地用一个阻值等于 k 的简单电阻 R 来终断电路。

这个网络假设由一个阻值为 k 的 Thevenin 电阻的信号源所驱动。注意在该例子中使用了两个完整的 T 形单元(在边框内)和在两端各自有半单元。(或者可以认为该结构由三个级联的 π 单元组成。)用半单元作为终断匹配是构架这类滤波器的传统方法。串联的电感，在图中单独地表示是为了清楚地表明每个 T 形单元的单独的贡献，在实际中则被组合成一个电感。相同的结构也可以用 π 形单元来构成(同样的，可以认为滤波器由三个完整的 T 形单元相级联)，如图 22.3 所示。

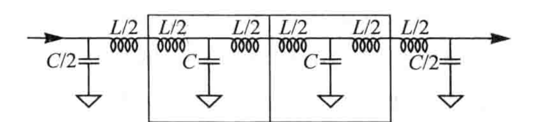


图 22.2　由两个级联 T 形单元构成的低通恒 k 滤波器例子

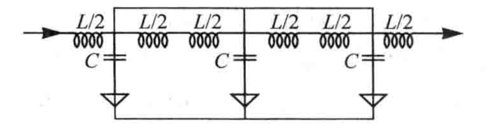


图 22.3　由两个级联 π 形单元构成的低通恒 k 滤波器例子

结构上的选择有时是基于要用到的电感的数目(在这两个例子中，是 3 对 4)，或者基于哪种拓扑结构能更好地容纳输出输入接口的寄生参数。如果寄生参数主要是容性的，那么 T 形单元更具有优势，因为寄生电容能够被端口处的电容所吸收。类似地，电感寄生效应能够被利用 π 形单元的滤波器更好地吸收。

表 22.1　理想恒 k 滤波器特性

n_s	在截止频率下的衰减/dB	归一化的 −3-dB 带宽	归一化的 −6-dB 带宽	归一化的 −60-dB 带宽	归一化的 −10-dB S_{11} 带宽
0	3.0	1.000	1.201	10.000	0.693
1	7.0	0.911	0.980	3.050	0.810
2	10.0	0.934	0.963	1.887	0.695
3	12.3	0.954	0.969	1.486	0.773
4	14.2	0.967	0.976	1.302	0.696
5	15.7	0.976	0.981	1.203	0.756

综合公式(3)与公式(4)，对于两种滤波器拓扑结构的设计公式可以容易地推导出，分别为

$$C = \left(\frac{2}{\omega_h}\right)\frac{1}{R} \tag{5}$$

$$L = \left(\frac{2}{\omega_h}\right)R \tag{6}$$

因此，一旦特征阻抗 R、要求的截止频率和单元的总数目确定了，滤波器的设计就完成了。

遗憾的是，推出单元的数目在实际中有些经验尝试的色彩。存在着一些公式，它们可以提供指导，但是它们往往或是太繁琐或是不够精确，使得通常是通过不断增加单元数目，直到仿真揭示了达到希望的滤波器要求[13]。而且，过于简单的用电阻来作为匹配将带来滤波器重要特性的变坏，通常的后果包括造成难以预测的插入损失和通带平坦度以及阻带衰减程度的减小(这是与对无限长传输线预测相比较)。这些困难从表格 22.1 中显而易见，表格表示了常数 k 滤波器在截止频率以及 −3dB、−6dB 的带宽(用截止频率的分数来表示)与阶数相关的衰减。在表中，n_s 表示完整的 T 形单元(或 π 形)在滤波器核心中的数目。滤波器的整体阶数 n 为 $2n_s+3$。

注意，带宽和在标称截止频率上的衰减都依赖于滤波器单元的数目。进一步注意到，截止频率(由公式(4)给出)仅仅在 $n_s=0$ 的情况下才等于 −3dB 带宽，而在最差的情况下甚至大于 −3dB 带宽 10%。在一些关键应用中，截止频率的设置或许需要根据带宽的要求而相应地作改变。

图 22.4 给出了由 5 个完整单元组成、在两终端由两个半单元匹配的常数 k 滤波器的频率响应图(根据运行 Puff 程序的结果)。注意频率轴是线性坐标而不是对数坐标。除了图中明显的波纹特性之外，令人沮丧的是回波损耗超过 10dB 的频率带宽仅是截止频率的 70%～80%。对于阻抗匹配带宽的一个相当大的改进可以通过使用这样的滤波器单元，其阻抗特性能更好地在一个更宽的归一化频率范围内近似常数电阻特性。由 Zobel 开发的一个例子使用了"m 导出式"的网络或当作匹配结构，或当作滤波器单元(或者两者都使用该网络)。在图 22.5 中，可以将结构看作是三个级联的 π 形单元或两个级联的 T 形单元(后者还有两个用于匹配的半单元)。

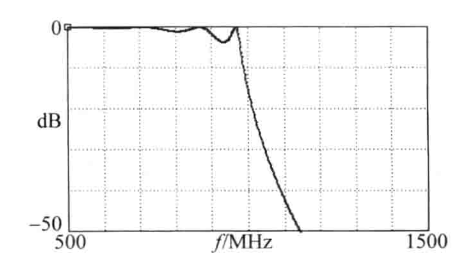


图 22.4　1GHz 低通常数 k 滤波器($n_s=5$)的频率响应

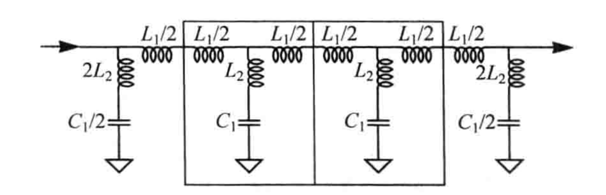


图 22.5　采用两个级联 T 单元的低通 m 导出式滤波器

[13] 比如，一个经常被引用的公式是基于一根无穷长梯形传输线的衰减特性。很明显，这样的一个公式，尽管十分简单，不可能被指望来产生对一根有限长、用电阻来终断的结构的有精确的预测。

与常数 k 滤波器的原型相同，这种结构的驱动和匹配都是一个阻值为 k 的电阻。这个 m 导出式滤波器，其本身就是一个常数 k 结构，能够通过以下的方式很好地被理解，这就是前面讨论的原型常数 k 滤波器在达到截止频率 ω_1 时通常衰减的十分剧烈。当频率是截止频率值的一小部分时，频率响应是比较平坦的，因此将截止频率增加到某个 ω_2 值应该在原先 ω_1 的带宽内产生更加平坦的频率响应，这一点看起来是合理的。

因此设计一个 m 导出式滤波器中的第一步是包括简单地增加原型常数 k 滤波器的截止频率。在不用 L_2 的情况下，我们可以看到用因子 m(m 在 0～1 之间)，去除 L_1 与 C_1，就会使截止频率按照因数 $1/m$ 而改变，从 ω_1 变为 $\omega_2=\omega_1/m$。而特征阻抗则保持 k 不改变，因为 L_1 与 C_1 的比例 k 并不因等比例缩减而改变。

现在，为了恢复原来的截止频率，添加一个电感 L_2 来产生与 C 的串联共振。在共振频率下，这个串联支路短路，导致在滤波器传输特性中出现一个陷波频率。如果这个陷波频率被置于一个合适的位置(略高于所希望的截止频率)，这个滤波器的截止频率可以被带回到 ω_1。然而需要注意，滤波器的频率响应在共振频率之上会上升回来(共振支路此时像一个简单的电感)。这个特性在研究 m 导出式滤波器以及与其类似的版本时，如 Chebyshev 和椭圆型滤波器(我们马上就会学习)，都要加以考虑。

遵循与决定普通常数 k 滤波器截止频率完全一样的步骤，我们发现 m 导出式滤波器的截止频率可以表示为

$$\omega_1 = \frac{2(R/L_1)}{\sqrt{4(L_2/L_1)+1}} \tag{7}$$

为了从公式中消去 L_1，注意到截止频率也可以表示成

$$\omega_1 = 2m/\sqrt{L_1C_1} \tag{8}$$

而特征阻抗由下式给出

$$R = \sqrt{L_1/C_1} \tag{9}$$

结合上述三个方程，L_2 可以解为

$$L_2 = \frac{(1-m^2)R}{2m\omega_1} \tag{10}$$

求解公式(8)与(9)，得到 L_1 与 C_1 为

$$C_1 = \left(\frac{2m}{\omega_1}\right)\frac{1}{R} \tag{11}$$

$$L_1 = \left(\frac{2m}{\omega_1}\right)R \tag{12}$$

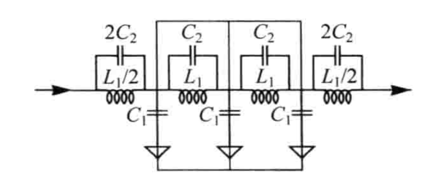


图 22.6 采用两个级联的带有并联谐振器的 π 形单元的低通常数 k 滤波器的例子

取代每一个滤波器单元中串联共振器的方案是采用并联共振器，如图 22.6 所示。L_1 与 C_1 的计算公式与之前相同：

$$C_1 = \left(\frac{2m}{\omega_1}\right)\frac{1}{R} \tag{13}$$

$$L_1 = \left(\frac{2m}{\omega_1}\right)R \tag{14}$$

C_2 的公式如下

$$C_2 = \frac{1-m^2}{2m\omega_1 R} \tag{15}$$

两种 m 导出式滤波器行为都相同。与常数 k 滤波器相同,在实际中拓扑结构的选择经常是由下面这个标准来决定:哪一种实现方式采用更容易或方便地实现的元件,或哪一种方式能够更好地吸收占主导定位的寄生元件。

使用前面的那些公式要求设计者对什么样的 m 值的选择有一定想法。当 m 趋近于 1 时,电路蜕变为一个通常的常数 k 滤波器(因此呈现出一个增加了的通带误差),而通带峰值随 m 接近 0 是增加的。通常在两种行为间取折中,实际中通常 m 在 0.5 附近的 25%～30%左右,最常见的选择是 0.6。后面这个数值在一个相当宽的频率范围之内基本可以保持传输系数的幅值为常数。表 22.2 列举了(表中有效位数要比实际需要的要多)$m=0.6$ 时的 m 导出式滤波器的一些相关特征。如表 22.1 所示,参数 n_s 为滤波器中使用到的完整的 T 形单元(或 π 形)数目。标有"最小阻带衰减"这一列给出了在那个传输陷波频率之上衰减的最坏值,在这个频率区域滤波器的响应又回到原来的水平。

表 22.2　理想的 m 导出式滤波器特性($m=0.6$)

n_s	在截止频率下的衰减/dB	归一化的 −3-dB 带宽	归一化的 −6-dB 带宽	归一化的 −10-dB S_{11} 带宽	最小阻带衰减/dB
0	1.34	1.031	1.063	0.965	8.21
1	3.87	0.993	1.013	0.956	21.24
2	6.27	0.988	0.999	0.969	34.25
3	8.30	0.989	0.996	0.979	47.09
4	10.00	0.991	0.995	0.954	59.81
5	11.44	0.993	0.996	0.961	72.43

注意,在表中截止频率和−3dB 带宽互相间要比常数 k 原型的情况更加接近(这里最差情况下的差别是 3%)。回波损失超过 10dB 的带宽占截止频率的比率也更高(事实上超过了 95%)。同样地从表中可以看到,在这个频率范围内,最小阻带衰减(以 dB 来衡量)大致等于 $8+13n_s$,因而可以通过对阻带衰减的要求轻松得到对单元数的估计。

图 22.7 显示了 m 导出式滤波器是如何提高频率响应的(注意纵坐标现在覆盖了 80dB,而不仅仅是 50dB)。相比较于图 22.4,这样的响应有明显小的通带波动,以及到阻带的更加陡峭的过渡(这是由于存在着阻带陷波点)。

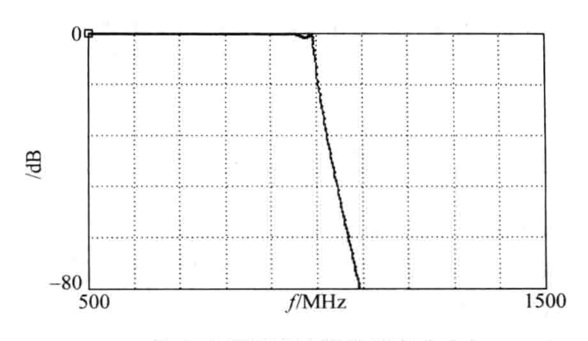


图 22.7　1GHz m 导出式低通滤波器的频率响应($m=0.6$,$n_s=5$)

在图 22.7 所示的频率范围尺度中，看不到带有特征性的陷波频响特性，同时也看不到在更高频率处的频响的再次上升。而这些特点在图 22.8 中被更加清晰地呈现出来。除了具有在通带内带来更加平坦的响应，m 导出式滤波器固有的陷波频率响应特性可以被用来消除一个特定频率上的干扰信号（或者是在一些频率处，如果采用了 m 值不相同的见过单元的话）。后面，我们会看到合理分布的陷波在 Chebyshev 和椭圆形滤波器中都有应用，使得通带到阻带的过渡更加陡峭。

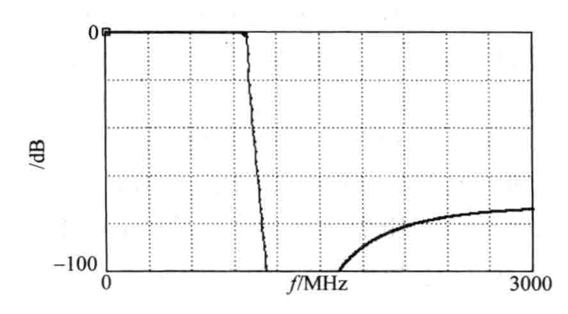


图 22.8　1GHz m 导出式低通滤波器在更广频率范围内的频率响应（$m=0.6$，$n_s=5$）

如果陷波的精确位置是重要的。对我们分析有益的是零点频率 ω_∞ 与 m 的关系，如下所示：

$$\frac{\omega_\infty}{\omega_1}=\frac{1}{\sqrt{1-m^2}} \tag{16}$$

因而为了在特定的频率 ω_∞ 上形成陷波，所需要的 m 为

$$m=\sqrt{1-\left(\frac{\omega_1}{\omega_\infty}\right)^2} \tag{17}$$

$m=0.6$ 将会使得陷波频率达到截止频率的 1.25 倍。

表 22.3 总结了对常数 k 与 m 导出式低通滤波器的设计。元件的值（同样地，比实际所需更加精确）是对应于 50Ω 电阻的终端（以及源）匹配与 1GHz 的截止频率。左侧的两列对应简单的常数 k 情况，而后面的四列给出了两种 m 导出式配置（针对 $m=0.6$ 的情况）的结果。对于截止频率高于 1GHz 的滤波器，可以简单地把所有的元件值乘以 1GHz 与希望达到的截止频率的比例。对于不同的特征阻抗，将所有元件的阻抗乘以希望达到的阻抗与 50Ω 的比值。

表 22.3　1GHz 常数 k 与 m 导出式滤波器的元件值（$Z=50\Omega$，$m=0.6$）

常数 k		m 导出			
L	C	L_1	C_1	L_2	C_2
15.9155nH	6.3662pF	9.5493nH	3.8197pF	4.2441nH	1.6976pF

当然也可以将常规的常数 k 与 m 导出式单元结合起来，因为每个单独的单元本质上都是恒定 k 型的。这样的混合滤波器在某些方面具有优势，例如，用于平衡通带的平坦度与陷波点出现在某些特定的频率。遗憾的是，这类滤波器的设计在很大程度上说是先验性的。简单地将单元混合和匹配，使得看起来可行，之后通过模拟来验证设计是否能够令人满意。

高通，带通和带阻形状

至少在理论上，一个高通常数 k 滤波器能够通过将在一个低通常数 k 滤波器原型中的电感和电容的位置互换就可以得到，而元件参数值不需要变。因此，用表 22.3 中的数值可以先设计一个 1GHz 的低通常数 k 滤波器，然后交换 L 与 C 综合得到 1GHz 的高通常数 k 滤波器。

加上"至少在理论上"这个限制词的原因是，高通滤波器的性能通常会与所期望的特性相距甚远。这个差距通常使得微波滤波器的设计者去避免使用基于低通原型得到的高通滤波器。尽管有很多——事实上是太多——的地方一个任何一类的实际滤波器的性能得不到预期的目标，接下来这个集总元件高通滤波器例子是足以说明这个问题的一般性质的。特别地，我们考虑图 22.9。

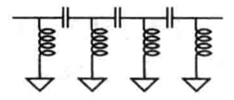


图 22.9　高通滤波器

每一个实际的电感会被一些寄生电容所旁路，因而呈现出其本身的共振特性。高于共振频率时，"电感"实际上将表现出电容的特性。类似的，每个实际的电容有某些与其串联的寄生电感。高于对应的串联共振频率，电容也表现出了电感的特性。因此，在足够高的频率下，一个高通滤波器的行为可能表现为低通滤波器的特性。其互补效应则可能影响低通滤波器，使得在高频区域的频响再次回升。

在需要重新设计的场合，采用一个通带足够宽的带通滤波器来近似所希望得到的滤波器特性的做法有时是有用的。当然，这种做法是预先假设了我们知道如何来搭建一个带通滤波器。幸运的是，常数 k 结构在这里也是适用的(稍后会讨论其他的通带实现方式)。作为一个从低通原型推导出带通滤波器一种通用策略，可以将一个低通原型的电感用 LC 串联来替代，同时将电容用 LC 并联替代。这些被添加的元件与已存在的元件在中心频率上共振，参见图 22.10。与我们之前的图不同，单独的 T 形单元并没有被显示出来(这是为了简化原理图)。

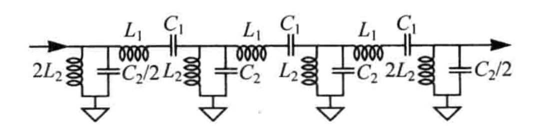


图 22.10　采用两个级联的 T 形单元的带通常数 k 滤波器例子

注意该结构继续呈现出正确的定性的行为，即使当电感最终呈现出电容特性，而反之时。这种性质对于潜在地减少拓扑结构对寄生效应的敏感度是带有根本性帮助的。

得到串联共振器中的电感 L_1 的公式与求低通原型滤波器电感是相同的，只是用带宽(定义为高低截止频率之差)替代了截止频率。电容 C_1 然后被选来在中心频率(定义为两个截止频率的几何平均值[14])处产生一个串联共振。因此

$$L_1 = \frac{2}{\omega_2 - \omega_1} R \tag{18}$$

$$C_1 = \frac{\omega_2 - \omega_1}{2\omega_0^2} \frac{1}{R} \tag{19}$$

[14] 在某些文献中，究竟应该用哪一种平均值这一点很遗憾地没有被表达清楚。对于通常的小的分数带宽，这个含糊是可以接受的，因为对于用算术还是几何平均值，差别并不大。实际元件的容差更使得这种微小的误差不重要。然而，这个差别随着分数带宽的增加而变大了，而且对大的分数带宽采用人工算术平均值，这个误差会变得十分显著。

类似地，要将低通原型中电容的公式变为对带通的情况，截止频率需要为带宽所替代。谐振电感再次被选来在中心频率处产生共振：

$$C_2 = \frac{2}{\omega_2 - \omega_1} \frac{1}{R} \tag{20}$$

$$L_2 = \frac{\omega_2 - \omega_1}{2\omega_0^2} R \tag{21}$$

对于一个截止频率为950MHz和1.05GHz的常数 k 带通滤波器(对应的中心频率近似为998.75MHz)的元件参数在表22.4中给出。与其对应的低通情况一样，半单元结构被用来进行带通滤波器的终端匹配。每个半单元由具有 $L_1/2$，$2C_1$，$2L_2$ 和 $C_2/2$ 值的元件组成。这样得到的滤波器具有与表22.1中列举的相同特性，这是假设比较是基于带宽而不是中心频率。

表 22.4　具有100MHz带宽，常数 k，1GHz带通滤波器的元件参数值($Z=50\Omega$)

L_1	C_1	L_2	C_2
159.15nH	0.15 955pF	0.398 88nH	63.662pF

作为一个特定的例子，考虑一个由 $n_s=5$ 的常数 k 原型导出的带通滤波器的频率响应(图22.11)。设计的带宽为100MHz，中心频率为1GHz。与期望相同，带通滤波器两边边沿处的特性和低通原型类似。

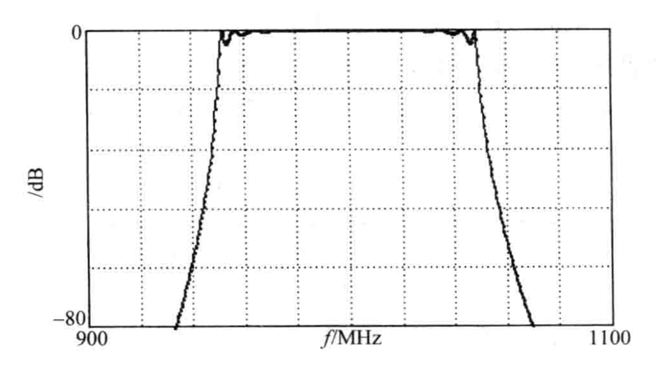


图 22.11　由常数 k 原型($n_s=5$)导出的带通滤波器的频率响应

对于不同的带宽，将 C_1 与 L_2 乘以新带宽与100MHz的比例，同时 L_1 与 C_2 降低相同的比例。对于不同的中心频率，按照新中心频率与1GHz比例的平方分别减小 C_1 与 L_2。最后，对于不同的特征阻抗，将所有四个元件的阻抗按照新特征阻抗与50Ω的比例增加。

带通滤波器可以通过交换串联和并联共振器的位置来转化成带阻滤波器(也可称作带排滤波器)。与低通到高通的转化相同，元件参数并不变。

从以上给出的表和例子可以清楚地看出常数 k 与 m 导出式滤波器设计起来非常简单，这是因为它们由相同的重复单元构成。这个简单的特点是它的最大优势。然而，如前所述，这种设计上的便利是以没法具体确定细节(如通带的波动)为代价的，因为这个设计方法并没有包括任何对频率响应形状的具体约束。例如，从那些表格中显而易见，截止频率并不对应于某个固定的衰减数值(如−6dB)，单调性远不能说是被保证了的。带阻行为同样没法

被控制。类似的这些缺点正是促使现代滤波器设计方法进步的动力。那些现代的方法允许我们在更多的细节方向上改造频率响应来满足更多设计实现上的要求。作为一个代价，这样的能力迫使我们要花一些时间来确定与定义那些滤波器性能的核心指标，以使得我们现在可以去规定。因而我们现在考虑一个简短的补充，并介绍这些参数。

22.4　滤波器的分类与指标

滤波器可以广泛地按照它们频率响应的形状来分类，例如，低通、带通、带阻与高通，也可以按照带宽、形状系数(或裙边选择性)以及抖动的程度(可以依据相位或者幅度的响应，以及在通带或阻带，或两者中)进一步更细致的分类。这类的细分是由于理想的、陡峭的滤波器特性是无法实现的(不仅是不实际的)。这些对理想特性近似的不同的方法导致了不同的折中，而这些折中的后果需要来进行表征。

带宽或许是一个最基本的描述性的参数，通常按照频响－3dB 点来定义。然而，需要指出的是 3dB 是一个相对随意的选择(能量减半点毕竟没有更基本的意义)，我们会依据具体所需，提出更加适合具体情况的带宽定义。这当然不能唯一确定滤波器设计，因为无限多的滤波器频响会具有相同的－3dB 带宽。形状系数能够在－3dB 点之外的频率给出一些滤波器频响特征，定义为两个不同衰减值之间的比例(如，在裙边区上两个不同点的数值)。作为一个特定的例子，形状系数为“6/60”定义为衰减值为－60dB 的带宽值除以衰减为－6dB 的带宽值得到的比例，见图 22.12。

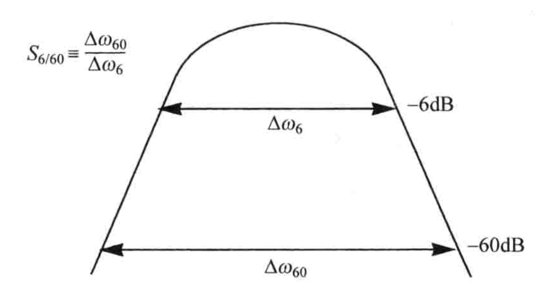


图 22.12　6/60 形状因子的示意图

从形状系数的定义中可以看到，数值趋于 1 明显对应着这样的频响形状，其从通带到阻带的过渡趋于无限陡峭。一个单极点低通滤波器(或一个标准单个 *LC* 带通谐振器)的 6/60 形状系数大约为 600，这通常被认为足够大[15]。这三个值是很容易被记住的，因为它们逐个增加一个零。因为一个给定的频响形状因子十分依赖于被应用的具体场合，因而无法给出一个普遍适用的定义。因此尽管 6/60 是一个较为常用的数值，其他的规定也经常会遇到。

如前所述，实际的滤波器无法实现完美平坦的通带与从通带到无限衰减阻带理想陡峭的过渡，这一事实指出我们必须接受对理想情况的合理近似。在最好的情况下，我们会有机会将近似误差的限定量化及确定。一个传统的做法是确定图 22.13 中所示的参数。注意图中画的是幅度的平方而不是幅度本身，这是因为前者与功率增益成正比。这个惯常的做法

[15] 实际的值是接近于 577，但这个值不如 600 来得好记。

并非都被采用，但是由于射频工程师所具有的典型的对功率增益的感谢，这个做法是十分常见的。

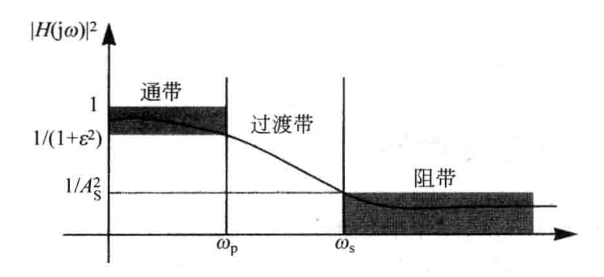


图 22.13　通用的滤波器响应的模板（图示的是对低通滤波器而言的）

同样可以注意到在纵坐标上常见的量的倒数形式。这个烦人的做法可以通过画出衰减（而不是增益）作为频率的函数来避免，这也解释了为什么许多处理方法用这样的形式来呈现数据。

进一步观察到滤波器频响的模板允许在通带内有一定量的变动（其上限被记为 ω_p），其波动的最大值为 $1/(1+\varepsilon^2)$。另外，从通带到阻带（其下限被记为 ω_s）有限的过渡也是允许的，其在阻带中允许的最小功率衰减为 A_s^2 这些参数的规定因此能允许实际滤波器的设计。现在考虑以这些参数为基础的几类重要的近似。

22.5　通用滤波器近似

常数 k 滤波器的那些局限性本质上是源自一个并不是综合而被称为综合的过程。它忽略了通过直接改变极点（和零点）的位置对滤波器频率响应的控制以达到要求的性能。这个局限性是常数 k 滤波器所基于的传输线理论基础而带来的一个自然结果，因为传输线是无限阶系统，尝试考虑极点的位置是不自然的，并且导致数学上的很大的困难。

然而，如果不再从传输线的角度来处理滤波器，那么这些就不再是问题（但是会被另外的新问题所取代）。其他一些功能强大的技术可以被应用到滤波器分析和综合的问题上。在这一节里，通过以一个并不是从恒的 k 滤波器的步骤来强调这一点：从希望得到的频率响应指标出发，计算一个相应的零极点图而后综合一个集总元件的网络，它在那些位置上有极点和零点。

我们在这个部分将要讨论的滤波器类型的频率响应，其幅度特征可以用下面的一般方式表示：

$$|H(\mathrm{j}\omega)|^2 = \frac{1}{1+\varepsilon^2 F^2(\mathrm{j}\omega)} \tag{22}$$

不同类型的滤波器依特定形式的 F 来区分。在最简单的情形下，F 是一个多项式，表示整个滤波器的传输函数仅包含有极点。对于更复杂的滤波器，该函数是多项式的比（即有理函数），允许传输函数也有有限个零点。如我们将要看到的，有理函数提供了比阶数 n 以外还要多的自由度，这样就使得我们能够降低所需要的滤波器阶数来满足一套给定的性能指标。

22.5.1 Butterworth 滤波器

在一些应用中完全不允许有抖动，这样就限制了频响形状的选择。如所有实际的滤波器一样，Butterworth 尝试去近似理想的矩形砖墙形状。Butterworth 滤波器的单调频率响应通过使在零频邻域中更多的不同阶导数为零以达到零频邻域的近似误差最小。当滤波器阶数接近无穷时，滤波器形状逐渐接近理想的矩形砖墙形状。

将更多的权重置于低频时近似误差的这样的设计理念一个自然(但潜在上是不希望的)的结果是误差在截止频率附近增加了。如果不希望看到这样的特征，应当寻找 Butterworth 之外的形式。接下来的节会讨论一些其他的选择[16]。

Butterworth 频率响应的幅度(平方)作为一个频率的函数在低通情况下，可以用下面的公式表示：

$$|H(\mathrm{j}\omega)|^2 = \frac{1}{1+\varepsilon^2\ (\omega/\omega_\mathrm{p})^{2n}} \tag{23}$$

其中在通带边频率 ω_p 处，功率衰减为 $1+\varepsilon^2$[17]。参数 n 为滤波器的阶(或度)数，等于独立储能元件的个数，这个参数也是 ω 的阶数，频响的幅度最终以这个行为滚降。很容易验证可以在 DC 时将导数值设为零的导数个数为 $2n-1$。从这个公式可以直接得到频率响应的确是单调的。

在设计一个 Butterworth 滤波器时，经常要给出 3dB 衰减频率 ω_c，这个频率(取决于允许的通带抖动)可能等于也可能不等于通带的边界[18]。为了保持与图 22.13 中模板的一致性，首先将 3dB 频率处功率增益表示为

$$\frac{1}{1+\varepsilon^2\ (\omega_\mathrm{c}/\omega_\mathrm{p})^{2n}} = \frac{1}{2} \tag{24}$$

从中容易决定

$$\omega_\mathrm{p} = \varepsilon\omega_\mathrm{c} \tag{25}$$

下一步，我们计算所需要的滤波器阶数，这是通过采用在阻带边界的衰减公式：

$$\frac{1}{A_\mathrm{s}^2} = \frac{1}{1+\varepsilon^2\ (\omega_\mathrm{c}/\omega_\mathrm{p})^{2n}} \tag{26}$$

求解需要的滤波器阶数 n，得到

$$n = \frac{\ln\left(\sqrt{A_\mathrm{s}^2-1}/\varepsilon\right)}{\ln(\omega_\mathrm{s}/\omega_\mathrm{p})} \approx \frac{\ln(A_\mathrm{s}/\varepsilon)}{\ln(\omega_\mathrm{s}/\omega_\mathrm{p})} \tag{27}$$

如果阻带衰减的平方比 1 大(通常是如此)，这个近似的效果较好，一旦最大通带衰减、最小阻带衰减和归一化的阻带频率确定，即可以计算得滤波器需要达到的阶数。由于公式(27)通常给出的不是整数，可以选择比该数大的整数作为滤波器阶数。在那种情形下，得到的滤波器会呈现出比原始预想的要更好的特性。利用这种“额外”的性能的一个方法是保留初始

---

[16] 就如在嗣后要讨论的，其中一个替代办法，逆 Chebyshev 或称第Ⅱ类 Chebyshev，通过允许在阻带中存在着抖动(同时保持通带内的单调性)而实际上取得比 Butterworth 更平的通带。

[17] 尽管并不是完全的严格正确(因为可能存在着不相等的输入与输出阻抗)，我们会频繁地交互使用“功率增益”与更为拗口的“响应幅度平方”。

[18] 许多关于 Butterworth 滤波器的讨论只考虑 $\varepsilon=1$ 的特殊情形，这相应于 $\omega_\mathrm{c}=\omega_\mathrm{p}$。而这里是考虑一般的情形，因此当把我们的结果与在其他地方发表的结果相比较时，不要发生混淆。

的 ω_p，如此滤波器在 ω_s 将给出更大的衰减。或者也可以保留原先的 ω_s，在这种情形下，滤波器在通带边缘表现出比预期更小的衰减(亦即更小的误差)。我们还可以选用一个在两者之间平衡的策略。

由于在直流处的近似误差非常小，Butterworth 通常被描述为具有最平特性[19]。然而，重要的是需要认识到最大化平坦并不代表完全平坦[20]。实际上，这表示在单调性的限制下能够达到的最平坦的通带(稍后，我们会看到如果允许在阻带里有抖动，甚至可以得到更加平坦的通带响应)。

作为一个设计例子，让我们继续用常数 k 滤波器拓扑结构这个例子。现在，与之前的例子相比，我们能够确定更多滤波器参数，这正是我们将要做的。这里，任意地假设在 1GHz 的通带边沿允许 1dB 的损失(即增益为 0.794)，并且在阻带边沿 3GHz 处，要求有 30dB 最小衰减。

从通带的要求，计算得 ε 为

$$\varepsilon = \sqrt{10^{(1\text{dB})/10} - 1} \approx 0.5088 \tag{28}$$

从阻带的要求特性，我们得到 A^2 为 1000。因此为了满足这个性能指标，所要求的最小滤波器阶数为

$$n = \frac{\ln(\sqrt{999}/0.5088)}{\ln 3} \approx 3.76 \tag{29}$$

其中取比其大的最小整数为 4。在接下来的计算中，对 n 假设这个数值并选择通带边界为精确的 1GHz。再一次，额外的性能可能用其他方式来分配，但是在缺少其他约束信息的前提下，我们做了这个任意的选择以便可以继续讨论。

下一步是计算元件的值。利用基于 22.6 节中讨论内容的方法，可以推导出 n 阶低通 Butterworth 滤波器元件参数的下面这个公式，归一化到 1Ω 阻抗以及 1rad/s 通带边沿：

$$b_k = 2(\varepsilon^{1/n})\sin\left[\frac{(2k-1)\pi}{2n}\right] \tag{30}$$

其中 k 的范围从 1 到 n。

如同在那些常数 k 例子中，Butterworth 滤波器可以从一个并联电容或串联电感开始。前面这个公式对于这两个等效配置都能给出归一化的值。如果滤波器从一个并联电容开始，带有奇数 k 的元件为电容，如果滤波器从串联电感开始，那么它们为电感。

下面这些式子给出了反归一化之后计算得到的实际元件的数值

$$L_k = \frac{R}{\omega_p} b_k \tag{31}$$

$$C_k = \frac{1}{\omega_p R} b_k \tag{32}$$

---

[19] 这个词明显来自：V. D. Landon 在"Cascade Amplifiers with Maximal Flatness"(具有最大化平坦度的级联放大器)，*RCA Review*，v. 5，January 1941，pp. 347-362。这个词的命名实际上是 Butterworth 自己在"On the Theory of Filter Amplifiers"(论滤波器放大器的理论，*Wireless Engineer*，v. 7，October 1930，pp. 536-541)对这个课题的描述要晚十余年。尽管其他人早期也发表过类似的结果，Butterworth 与最大平坦化这两个词现在看起来是永远联系在一起了。

[20] 在某种程度上，"最大化平坦"有点像在描述 Oreo 这个牌子的点心的成分时用了"填满了奶油"这个说法。这意味着一些事情并不完全似它们被称呼时的那样。

使用这些公式得到如表 22.5 所示的元件值，在这个表中对两种等效实现的数值显示在并排的两列中。注意两种设计实际上是相同的，唯一的不同是输入和输出端口的规定。一般而言，偶数阶 Butterworth 滤波器共享这种特征，而奇数阶滤波器是对称的，即头尾为相同的元件(并且其值也相同)。

表 22.5　1GHz，1dB 的四阶低通 Butterworth 滤波器的元件值($Z=50\Omega$)

L first	C first
$L_1=5.1441\text{nH}$	$C_1=2.0576\text{pF}$
$C_2=4.9675\text{pF}$	$L_2=12.419\text{nH}$
$L_3=12.419\text{nH}$	$C_3=4.9675\text{pF}$
$C_4=2.0576\text{pF}$	$L_4=5.1441\text{nH}$

对任何一个设计进行仿真来确认没有计算误差的发生总是一个聪明的做法。注意，即便使用数据表中的值——一些发表的表格有排版上的错误！在一些要求比较高的应用中，为了评价滤波器对元件值变动或其他一些不理想因素(例如有限的元件的 Q 值或者寄生参数的存在)的敏感度，模拟也是十分有价值的。RFSim 99 这个程序值得提到，因为使用它可以快速地估计器件容度对滤波器频响的影响。尽管其他一些仿真器也提供 Monte Carlo (MC)仿真，RFSim 99 仅用很少的几个键击或鼠标点击就可以方便地调用 MC 仿真。图 22.14 显示了这个 Butterworth 滤波器对于器件参数公差的敏感度并不大(在这个情况下，电容和电感分别有 5%的允许公差，这对于一般器件而言是精度比较高的)。即便如此，仿真依然显示相当大一部分设计不能满足设计指标。需要或者是更严格的元件公差或更高阶的滤波器设计来确保高良品率同时允许更宽松的公差。

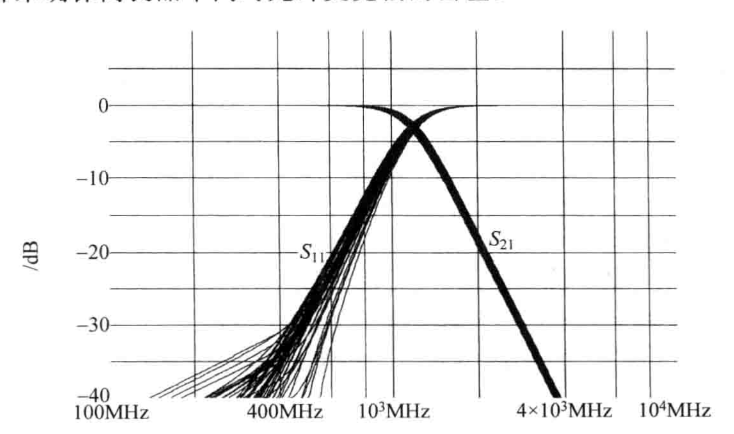


图 22.14　Butterworth 滤波器幅度响应的蒙特卡罗分析

最后，Butterworth 高通、带通和带阻滤波器可以通过与常数 k 情况相同的变换方式很方便地得以实现。

22.5.2　Chebyshev(等抖动或最小最大)滤波器

尽管通带和阻带的单调性看上去十分吸引人，坚持要保持这种性质将限制其他重要的滤波器特性。这些其他的特性包括从通带到阻带过渡的陡峭度，也包括对给定的滤波器阶

数的阻带衰减。其他基于单调频响的滤波器，是以发明它或者推导相关数学内容的人命名的。Chebyshev 滤波器是后者（数学推导）的一个例子，这种滤波器通过放宽单调性的限制来降低滤波器阶数[21]。与强调在低频段尽量降低误差的 Butterworth 近似相反，Chebyshev 要在整个通带上减小最大近似误差（这是相对于与理想的砖墙形状比较而言）。因此得到的最小化极大值的频率响应形状表现出一些抖动，而抖动的量可以由设计者控制。对于给定的阶数，Chebyshev 滤波器相比 Butterworth 能够提供从通带到阻带的更陡峭的过渡。这个过渡的陡峭程度也是设计者可允许的通带中的抖动大小的函数；能够允许的抖动越大，过渡就更为陡峭。

最小化最大误差的结果是 Chebyshev 频响的抖动的幅值都相同。一个对最小化最大值的优化方法产生抖动幅度上等同的完整的证明是非常繁琐的，我们在这里就不予推导。然而，在直觉上就可以理解到出于最小化极大值的考虑，等幅抖动应该是最优的。这是因为如果一个误差峰值大于其他误差峰，则更好的近似可以通过减小这个峰值来达到，这又是以增大了一个或其他几个峰值为代价的。这样的折中将一直进行下去，直到没有可以再进行折中的可能，那时所有的误差峰值就都是相等的了。

对于阻带而不是通带，类似的优势同样可以得到，如果允许抖动的存在。逆 Chebyshev 滤波器（也被称作第二类 Chebyshev 滤波器）就是基于这样的主意，它实际上将一个比 Butterworth 更平的通带与一个等幅抖动的阻带结合在一起。

为了理解为何通过简单地允许阻带或通带有抖动就能够提供这些好处，需要回顾一下复极点对的一些性质。首先回想起对于这样极点对的传输函数的一个标准（并且是十分通用的）形式为

$$H(s)=\left[\frac{s^2}{\omega_n^2}+\frac{2\zeta s}{\omega_n}+1\right]^{-1} \tag{33}$$

其中 ω_n 为极点到原点的距离，而 ζ 是阻尼系数，参见图 22.15。

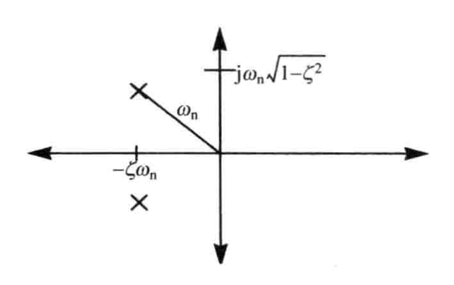


图 22.15　双极点星座图

阻尼系数对于研究时域特性是一个特别关键的参数。零阻尼对应于两个纯虚数极点，而阻尼系数为 1 对应于实轴上一对相重合的极点。前面这个条件适用于振荡器，而后者则定义了临界阻尼。超过了阻尼系数 1，两个极点在实轴上分裂：一个向原点移动，另外一个向负无穷移动。不论阻尼数值的大小，频率 ω_n 永远等于两个极点频率的几何平均值。

在滤波器设计中，我们通常最关心的是频率响应，因此直接将阻尼系数与那些更加直接相关的参数联系起来对分析是有帮助的。特别的是，阻尼系数与 Q 通过下面的公式直接

---

[21] 与通常的知识相反，Pafnuti L'vovich Chebyshev 从来就没有对滤波器进行过研究（他出生在 1821 年，去世于 1894 年，这要比滤波器理论的确立要早许久）。事实上，他在研究一个用在蒸汽机中一个机械连接时发展了他的这些方程，如在“Theory of Mechanisms Known under the Name of Parallelograms”（被称为平行图的机制的理论），*Oeuvres*, vol. 1, St. Petersburg, 1899。“Parallegrams”将旋转运动转化为直线运动。顺便说一下，他的名字在这里的拼法只是许多可能的翻译的一种译法。德文期刊通常将其写作 *Tchebescheff*。我们已经选择了这样的音译以使得能让英语的发音最接近于原来的发音。而通常遇到的 *Chebychev* 则缺乏这个性质。

相关：

$$Q = 1/2\zeta \tag{34}$$

尽管对于没有有限值零点的两极点系统其频率响应最终是以 ω^{-2} 的形式小滚降，这一点是成立的，频率响应的幅度与在峰值邻域内的斜率都为 Q 的强函数，两者都随 Q 的增加而增加(参考图 22.16 中的下方的子图)。对于小于 $1/\sqrt{2}$ 的 Q，频率响应没有峰值出现。超过这个值的 Q，随 Q 增长到无穷，峰值无上界的增长。对于较大的 Q，峰值的增益与 Q 成正比。换句话说，更高的 Q 值带来更大的最终衰减，这是相对峰值增益以及斜率而言的。这里的斜率通常是与更高(或许是高得多的)阶的系统相联系的。

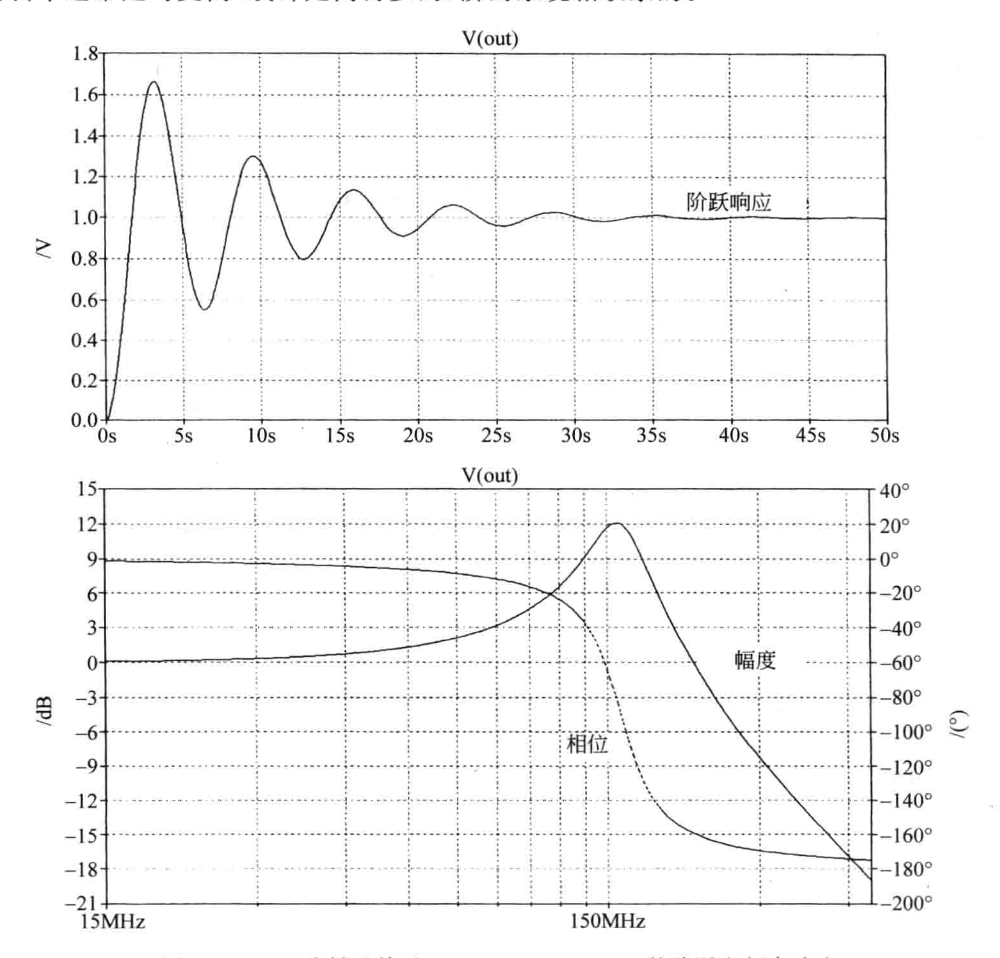


图 22.16　二阶低通单元($\omega_n = 1\text{rps}, \zeta = 0.125$)的阶跃和频率响应

现在我们考虑滤波器可能利用这种依赖 Q 行为的方法。特别的是，设想用一个二阶单元来提高一个单极点滤波器的频响幅度特性。如果在一阶单元响应开始有大的下降时，用二阶响应的峰值来补偿(即提升)其响应，那么这个级联结构的幅度大致保持常数的频率区间将会增加。同时，在补偿点之后的衰减将呈现一个相当高的初始斜率，从而提供了一个由通带到阻带的改善了的过渡。明显地，更多额外的单元可以被用来提供更大的提升，同时带来最大 Q 值的增加。后面这个要求源自对不断变大的衰减提供更大补偿的需要。

有了这样的理解，可以重新访问 Butterworth 和 Chebyshev 近似。Butterworth 条件的得到是通过调整转移特性的极点以使得一对复极点的一定的频响峰值能够在某种程度上抵消任何在更低频率极点的滚降。这样得到的组合增加了传输幅度基本常数的频率范围。形式上，可以从公式(23)中推断出所有极点都处于 s 平面的一个半圆上，它们的分布方式好像是有两倍数目的极点等角度地置于圆周上，然后将右平面的极点忽略掉[22]。比如说，一个三阶 Butterworth(图 22.17)有一个实轴上的单极点，和一对与实轴成 60°角的共轭复极点。从极点到原点的距离为−3dB 截止频率。

Chebyshev 滤波器更进一步允许通带(或阻带)有抖动。继续讨论我们这个三阶例子，这个实极点的响应被允许在那对复极点的峰值将响应拉回来之前掉在那个低频值之下的一定值(即被允许的抖动量)。复极点对的阻尼系数必须比 Butterworth 的相应的对要低，以产生足够的额外成峰来补偿更多的衰减。

这个降低了的阻尼系数的一个副作用是在截止频率之后的更为剧烈的衰减。正是以这样的方式，Chebyshev 滤波器允许设计者在通带的平坦度与更好的阻带衰减之间作折中。

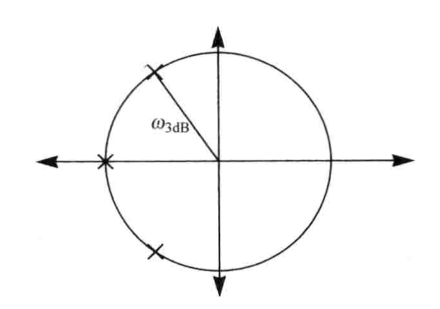


图 22.17　三阶 Butterworth 低通滤波器的极点星座图

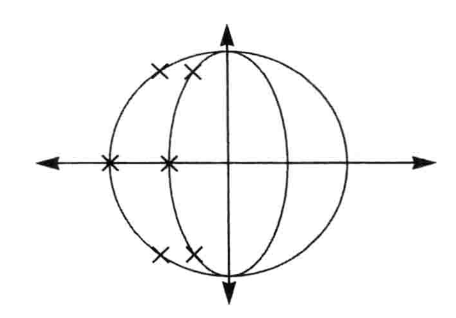


图 22.18　三阶 Butterworth 和 Chebyshev 低通滤波器的极点星座图

尽管在直觉上可能更不容易理解，Chebyshev 低通滤波器的极点都位于一个(半)椭圆上，尤其令人惊奇的是极点的虚部等于对应的 Butterworth 低通滤波器极点的虚部[23]。增加椭圆的偏心率会增加抖动，参见图 22.18。数学上，Chebyshev 响应具有以下的基本形式

$$|H(j\omega)|^2=\frac{1}{1+\varepsilon^2 C_n^2(\omega/\omega_p)} \tag{35}$$

其中 ω_p 同样还是那个频率，在该频率下响应的幅度的平方值下降到下面这个值

$$\frac{1}{1+\varepsilon^2} \tag{36}$$

出于不言自明的原因，ε 被称为抖动的参数，其值由设计者确定。函数 $C_n(x)$ 被称作阶数为

[22] OK，或许这并不是那么“直觉上是明显的”，但是求出方程(23)的所有的根以发现关于 Butterworth 极点位于一个圆圈上不是一件坏事。

[23] 有许多文献提供了极佳的 Butterworth 与 Chebyshev 条件的推导。一个特别说明问题的描述可以在 R. W. Hamming 的 *Digital Filters* (2nd ed.，Prentice-Hall，Englewood Cliffs，NJ，1983)(数字滤波器)的第 12 与 13 章中找到。

n 的 Chebyshev 多项式。该类多项式最相关的性质是当 x 在-1 与＋1 之间改变时，多项式的值在－1 与＋1 之间振荡。这个性质将近似所带来的负担通过允许滤波器的（功率）响应在通带内在 1 与 $1/(1+\varepsilon^2)$间振荡来平均地分散开来。回想一下，Butterworth 滤波器将所有的近似集中在直流点上，使得当通带边界接近时误差单调增加。Chebyshev 滤波器则通过将近似误差分散到整个通带上以取得更好的性能。在这个区间之外，Chebyshev 多项式的大小快速增加（事实上随 x^n 而增加），这对应于单调增加的滤波器衰减。

存在着一些用算法来产生 Chebyshev 多项式的方法。其中之一是通过下面的迭代

$$C_n(x) = 2xC_{n-1}(x) - C_{n-2}(x) \tag{37}$$

其中令 $C_0=1$ 和 $C_1=x$ 可以开始迭代过程。从公式中可以看到，Chebyshev 多项式的主要项系数为 2^{n-1}，这个事实是我们稍后比较 Chebyshev 和 Butterworth 多项式时要用到的。

另一种产生 Chebyshev 多项式的方法是通过某些三角函数，从这些函数中－1 与＋1 之间的振荡（对$|x|<1$）可以直接推出：

$$C_n(x) = \cos(n\arccos x), \quad |x|<1 \tag{38}$$

对于大于 1 的情况，公式略加改变为

$$C_n(x) = \cosh(n\operatorname{arccosh} x), \quad |x|>1$$

尽管此时看来完全不明显，读者将这些函数看作李萨如（Lissajous）图形会更加熟悉，这个图形是当用正弦波信号同时驱动示波器的水平和垂直偏转板时而形成的。亦即是，假设水平偏转板由以下信号驱动

$$x = \cos t \tag{40}$$

那么

$$t = \arccos x \tag{41}$$

进一步假设垂直偏转板同时由以下同步信号驱动

$$y = \cos nt \tag{42}$$

将公式（41）代入到公式（42）来消除时间参数，得到

$$y = \cos(n\arccos x) \tag{43}$$

这与公式（38）看起来是相同的。亦即，被如此驱动的示波器上显示的实际上就是具有那个 n 阶的 Chebyshev 多项式，$|x|$被限制在 1 之内。在这个区间内，所显示的函数看起来很像将正弦函数画在一张纸上，然后卷成一个圆柱体，再从很远的距离看过去。

一些 Chebyshev 多项式的简图粗略地画在图 22.19 上，前 10 个 Chebyshev 多项式的表达式在表 22.6 中列出。根据前面的这些公式，可以推导出一个为了满足给定的约束所需要的滤波阶数：

$$n = \frac{\operatorname{arcosh}\left(\sqrt{A_s^2-1}/\varepsilon\right)}{\operatorname{arcosh}(\omega_s/\omega_p)} \approx \frac{\operatorname{arcosh}(A_s/\varepsilon)}{\operatorname{arcosh}(\omega_s/\omega_p)} \tag{44}$$

这一组方程与用于 Butterworth 情形的方程在形式上是类似的，唯一的区别是自然对数被替代为 arcosh。同样的，如果阻带衰减的平方远大于 1，那么这个近似成立比较好。另外，与 Butterworth 相同，由公式（44）给出的阶数取到下一个整数。同时，得到“额外”的性能可以被用来提升某种通带和阻带的特性的组合。

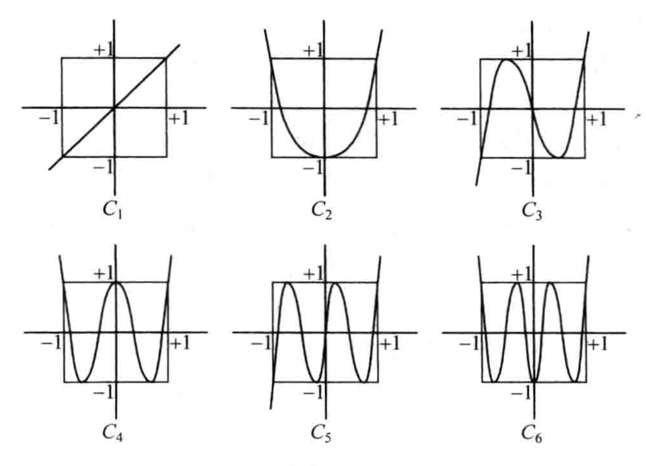


图 22.19 一些 Chebyshev 多项式的粗略草图

表 22.6 前 10 个 Chebyshev 多项式

阶数，n	多项式
0	1
1	x
2	$2x^2-1$
3	$4x^3-3x$
4	$8x^4-8x^2+1$
5	$16x^5-20x^3+5x$
6	$32x^6-48x^4+18x^2-1$
7	$64x^7-112x^5+56x^3-7x$
8	$128x^8-256x^6+160x^4-32x^2+1$
9	$256x^9-576x^7+432x^5-120x^3+9x$

Chebyshev 比 Butterworth 更好的一方面是前者可提供的最终阻带衰减。在高频范围，一个 $\varepsilon=1$ 的 Butterworth 滤波器能够提供的衰减大约是

$$\left|A\left(\mathrm{j}\frac{\omega}{\omega_\mathrm{p}}\right)\right|^2=\left(\frac{\omega}{\omega_\mathrm{p}}\right)^{2n} \tag{45}$$

将渐近行为与 Chebyshev 的比较（同样设 $\varepsilon=1$）[24]：

$$\left|A\left(\mathrm{j}\frac{\omega}{\omega_\mathrm{p}}\right)\right|^2\approx 2^{2n-2}\left(\frac{\omega}{\omega_\mathrm{p}}\right)^{2n} \tag{46}$$

明显的，对于给定的阶数，Chebyshev 滤波器提供一个高出 $3(2n-2)$dB 的最终的衰减。作为一个特定的例子，一个七阶 Chebyshev 滤波器比七阶 Butterworth 滤波器最终要提供 36dB 的更高的阻带衰减。

作为另外一个比较，通过将具有相同阶数的 Butterworth 与 Chebyshev 滤波器都归一

[24] 这个比较不应该引导你错误地认为这样大的抖动是通常用到的。事实上，这种滤波器典型地是设计成抖动值低于 1dB。

化到同样的 -3dB 带宽，这两个滤波器的极点的关系可以被置于一个定量的基础。也可以证明(但不是被我们来做)Chebyshev 的 -3dB 带宽可以被合理地近似为[25]

$$\cosh\left[\frac{1}{n}\operatorname{arsinh}\left(\frac{1}{\varepsilon}\right)\right] \tag{47}$$

由于 Butterworth 的极点圆形星座图的半径为 -3dB 带宽，将 Chebyshev 椭圆归一化以使得长轴如公式(47)定义。这些 Chebyshev 滤波器极点的虚部与 Butterworth 相同，而与 Butterworth 原型的实部仅用如下的因子来成比例缩放

$$\tanh\left[\frac{1}{n}\operatorname{arsinh}\left(\frac{1}{\varepsilon}\right)\right] \tag{48}$$

来得到 Chebyshev 滤波器极点的实部。因此 Chebyshev 滤波器的设计可以基于 Butterworth 原型，而后者的设计十分容易。

然而还有一点微妙之处需要讨论，而这是关系到无源 Chebyshev 滤波器源与终端的阻抗。从画的简图与公式中都可以看到，很明显只有奇数阶 Chebyshev 多项式在自变量取零时其值为零。因此，对于这样的多项式，滤波器传输函数在直流处为 1(亦即，通带的第一次掉到 1 以下发生在直流之上的某个频率)。然而，对于偶数阶 Chebyshev 滤波器，滤波器传输函数自一个下跌开始，而其直流的功率传输值为 $1/(1+\varepsilon^2)$，意味着终端匹配电阻小于源端电阻。如果，通常也正是如此，不希望有这样的阻抗转换，那么选择一个奇数阶的 Chebyshev 滤波器或对一个偶数阶的 Chebyshev 滤波器加一个阻抗转化器。因为前者比较不复杂，奇数阶的 Chebyshev 实现在实际中几乎是通用的选择。

最后，意识到椭圆形的极点分布意味着极点的虚部与实部的比例，因此也就是极点的 Q 值，Chebyshev 的比同阶的 Butterworth 要高。结果是，Chebyshev 滤波器在更大程度上受到实际元件的有限 Q 值的影响。这个问题的严重性随着滤波器阶数的增加而变得更为严峻。这个重要的实践因素在选择滤波器类型时必须要加以考虑。

考虑到 Chebyshev 与 Butterworth 的极点位置紧密相关，发现两种滤波器的元件数值也彼此相关这个事实，也不应该感到奇怪了。事实上，Chebyshev 滤波器元件的参数可以利用下面的一些公式由 Butterworth 推导得出[26]。首先计算一个参数 β，将下面这个表达式进行简化：

$$\beta = \sinh\left(\frac{\operatorname{arctanh}\left(1/\sqrt{1+\varepsilon^2}\right)}{n}\right) \tag{49}$$

那些元件值便为(同样归一化到 1rps 和 1Ω)

$$c_1 = b_1/\beta \tag{50}$$

与

$$c_k = \frac{b_k b_{k-1}}{c_{k-1}\left(\beta^2 + \left\{\sin\left[\frac{(k-1)\pi}{n}\right]\right\}^2\right)} \tag{51}$$

---

[25] 见，比如说，M. E. VanValkenburg, *Introduction to Modern Network Synthesis*(现代网络综合导论)(Wiley, New York, 1960), pp. 380-381。最早的方法是由 E. A. Guillemin 提出的。

[26] 读者可以用 22.6 节中给出的那些方法来推导出这些方程(或它们的等价公式)，但是这里给出的特别的方程(在变量的使用上有一些小的改动)摘自 David B. Rutledge 的出色教科书 *The Electronics of Radio*(无线电电子学)(Cambridge University Press, 1999)。

其中不同的 b_k 依然是归一化了的 Butterworth 元件值(对于 $\varepsilon=1$)：

$$b_k = 2\sin\left[\frac{(2n-1)\pi}{2n}\right] \tag{52}$$

值得指出的是，这些公式仅仅对 n 为奇数情况适用，因为它们假设了同样的源与负载终断。如前所述，偶数阶的 Chebyshev 滤波器要求不相等的源和负载终断(除非允许采用阻抗转换器)，因此相比较奇数阶并不常用[27]。这里列出的公式组因此能满足大部分应用的要求。

继续进行当前的滤波器设计例子，首先决定所需要的最小阶数

$$n = \frac{\operatorname{arcosh}(\sqrt{A_s^2-1}/\varepsilon)}{\operatorname{arcosh}(\omega_s/\omega_p)} \approx \frac{\operatorname{arcosh}(31.6/0.5088)}{\operatorname{arcosh}([3\text{GHz}]/[1\text{GHz}])} \approx \frac{4.8222}{1.7267} = 2.73 \tag{53}$$

因此，为了满足这些性能指标，三阶的 Chebyshev 就够了。与 Butterworth 的例子相同，我们将任意地选择首先来严格满足在通带的指标，而将额外的性能留给阻带。一个更为实际的选择是将额外的性能分配在两者之间，但是我们先就这样做。所得的三阶滤波器的标称元件数值如表 22.7 所示。

表 22.7　三阶 1dB-1GHz，30dB-3GHz Chebyshev 滤波器元件值

L first	C first
$L_1=16.104$nH	$C_1=6.442$pF
$C_2=3.164$pF	$L_2=7.911$nH
$L_3=16.104$nH	$C_3=6.442$pF

对于这个标称设计的频率响应如图 22.20 所示，元件的容差假设为 5%，滤波器传输特性的分布如图 22.21 分布。

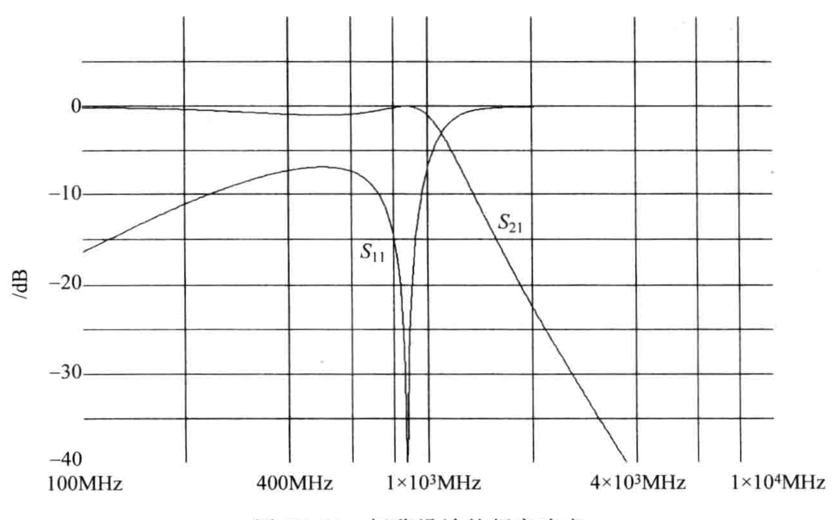


图 22.20　标称设计的频率响应

[27] 然而，注意到阻抗变换有时是希望有的，在这些情况下，偶数阶 Chebyshev 滤波器可能是十分有用的。

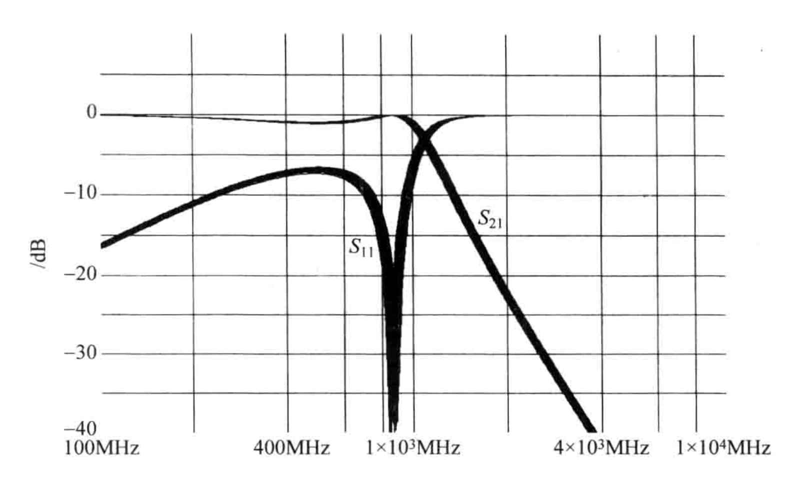


图 22.21　1-GHz，1-dB 三阶 Chebyshev 滤波器的蒙特卡罗仿真(抖动方差为 5%)

22.5.3　第二类(逆)Chebyshev 滤波器

我们已经几次提及了实现比极大化平坦度更加平坦的传输特性的可能性。第二类(也叫做逆或倒)Chebyshev 滤波器通过允许在阻带中存在抖动来取得这种平坦度，同时还保持了通带中的单调性。

第二类滤波器通过一对简单的变换从第一类(普通的)Chebyshev 滤波器得到。第一步，将 1 减去第一类 Chebyshev 的频率响应，从而实现了从低通滤波器到高通滤波器的转换。得到的频率响应在新的通带上是单调的，这是因为第一类滤波器在其阻带区是单调的。下一步所要做的就是找到一种方法将高通滤波器在保持单调性的前提下将其转化为低通滤波器。这第二步的关键是将 ω 替换为 1/ω。因为如此高频便对演为低频，反之亦然，这个第二步转换的确把滤波器的形状转换回低通响应，但是是通过一个将低频区域的抖动换到了高频区的抖动的方法。这个变换就这样恢复了通带的单调性，同时第一类通带的边沿也正好对演到了新的阻带边沿。这里的折中是用失去阻带单调性的代价换取了更平坦的通带。

数学上，上面所述的变换得到了下述对第二类滤波器的功率响应：

$$|H(\mathrm{j}\omega)|^2 = 1 - \frac{1}{1+\varepsilon^2 C_{\mathrm{n}}^2(\omega_{\mathrm{p}}/\omega)} = \frac{\varepsilon^2 C_{\mathrm{n}}^2(\omega_{\mathrm{p}}/\omega)}{1+\varepsilon^2 C_{\mathrm{n}}^2(\omega_{\mathrm{p}}/\omega)} \tag{54}$$

归一化了的逆 Chebyshev 滤波器因此有极点位于与正常的(第一类)Chebyshev 取逆的位置，此外还有一些以一种较为复杂的方法分布的纯虚数的零点。正如复极点对会带来峰值一样，一对复零点对会带来零响应点。我们已经看到过这样的特性，在那里 m 导出式滤波器中纯虚数零点带来无限深的陷波点。逆 Chebyshev 滤波器利用这些零值点来提供平坦的通带而不会使由通带到阻带的过渡变坏。得到的零极点星座图大致相似于逆时针旋转 90°后的希腊字母 Ω。

尽管第二类 Chebyshev 滤波器没有 Butterworth 那样常见，它的少见并不应误解为它用到的时候也不多。尽管由逆 Chebyshev 滤波器提供的极佳的通带平坦度，缺少对器件参数的简单公式使得 Butterworth 滤波器在对于通带单调性要求比较高的应用中占主导地位。幸运的是，滤波器设计软件的出现改变了这个状态。例如，程序 LADDER 能够综合很

多类无源滤波器，逆 Chebyshev 滤波器是其中之一。通过使用这一软件，我们得到图 22.22 中的低通滤波器（在 1GHz 处 1dB 的误差，在 3GHz 处至少 30dB 的衰减）。

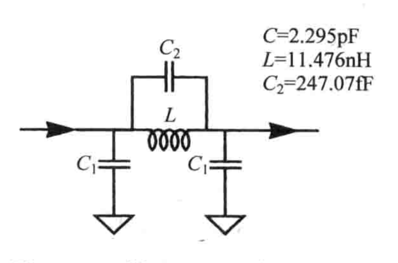


图 22.22　低通 1GHz 逆 Chebyshev 滤波器的例子

我们已经忽略了在两个端口处采用串联电感的对仗网络的综合，也忽略了在它们的公共点与地之间的串联 LC 陷阱。图 22.23 中显示了逆 Chebyshev 滤波器与 Butterworth 相比确实提供了从通带到阻带更快的过渡，同时保持了通带响应的单调性。与此同时，在通带内的平均回波损失也比普通的 Chebyshev 滤波器更大。

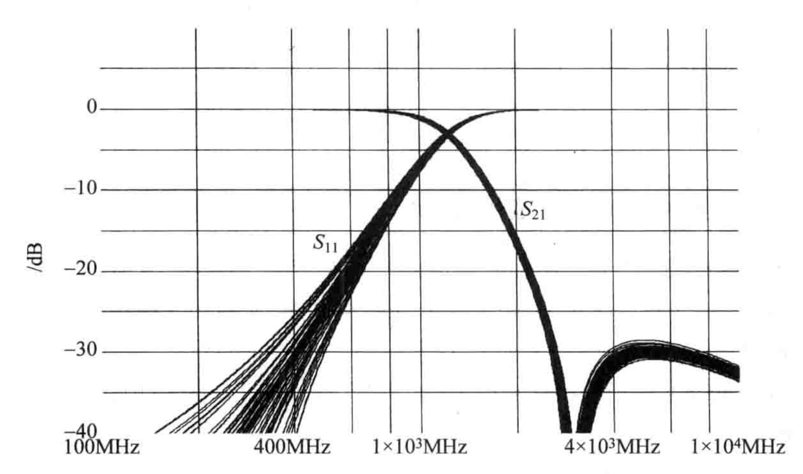


图 22.23　逆切比雪夫滤波器的蒙特卡罗模拟结果（抖动方差容差为 5%）

22.5.4　椭圆（考尔）滤波器

我们已经看到在通带或阻带允许抖动能够带来有用的特性，因此在阻带和通带同时允许抖动以改善在过渡过程的陡峭程度或许不是一件奇怪的事[28]。就如在逆 Chebyshev 与 m 导出式滤波器，考尔（也称作椭圆型）滤波器利用由有限个零点提供的陷波点来建立从通带到阻带的极快的过渡——这是以阻带响应会在陷波点之后的频率上反弹为代价的（这与逆 Chebyshev 与 m 导出式滤波器一样，源自同样的原因）。如在 22.7 节中要进一步讨论的，椭圆型这个名称来自数学上的椭圆方程的出现，而不应与 Chebyshev 滤波器极点的椭圆形分布相混淆。

Wilhelm Cauer 是该滤波器的发明者，他的深入的物理内在理解（以及对 Zobel 的 m 导出式滤波器的陷波点与更一般的椭圆函数的十分熟悉程度）使得他能首先意识到这些存在的额外的自由度，并且加以利用。尽管他首次发表椭圆型滤波器时，对他的想法没有完整的

---

[28] 这一些有时也被称为 Darlington，Cauer-Chebyshev，广义 Chebyshev 或 Zolotarev 滤波器。Igor Ivanovich Zolotarev（1847－1878），他曾经师从 Chebyshev，很明显在 Chebyshev 这样做十年左右之前就推导得到了 Chebyshev 函数。他的重要贡献包括对椭圆函数的数值近似及用有理函数来得到最小化极大值近似。在 1878，他也通过直接的实验演证了列车与人体间动量转移的悲惨后果（Zolotarev 于 1878 年因火车事故去世，译者注）。

数学证明[29]。在一个专注于减少元件数目的时代，Cauer 能够使用比当时最好的滤波器用到的更少数目的电感。根据记载，他的专利的发表使得 Bell 实验室的工程师和数学家涌向纽约市的公共图书馆好几周去学习当时（现在依然是如此）的关于椭圆函数的深奥的文献[30]。考虑到 Bell 实验室的那么聪明的学者都花了如此多的时间，我们没指望在这里能够将问题解释清楚。下面只给出一个大纲。

椭圆型滤波器有如下的功率传输函数特性：

$$|H(\mathrm{j}\omega)|^2=\frac{1}{1+\varepsilon^2F_{\mathrm{n}}^2(\omega/\omega_{\mathrm{p}})} \tag{55}$$

这里 $F_n(x)$ 是两个多项式的比而不只是一个简单的多项式（后面这个情况与 Butterworth 或 Chebyshev 相同）。一个关键的观察结果是，对于给定的阶数，这样一个有理函数相比较一个简单的多项式有更多的自由度（亦即有更多的系数可以调整用于达到设计的目标）。这些额外的自由度允许以更低的阶数来满足设计标准。因此，基本的动机并不是那么难掌握，但是细节才是关键的所在。

就像对于 Chebyshev 多项式那样，这些有理函数 $F_{\mathrm{n}}(x)$（被称作 Chebyshev 有理函数）的值（幅度）在变量 $|x|$ 小于 1 的情况在 0 与 +1 之间振荡，在变量的这个范围之外幅度则并迅速增长，而且阶数 n 增长时，函数的幅度也增长[31]。然而，与 Chebyshev 多项式其幅度在那个范围之外单调增长不同，$F_{\mathrm{n}}(x)$ 在无穷与一特定有限值之间以某种形式振荡。因此滤波器的频率响应呈现出在阻带内的抖动，而且在一些频率点上传输系数为零。那些衰减的极点对应于传输的零点（即陷波点）在其邻域内滤波器响应快速地发生变化。因此，或许读者可以理解在阻带内允许抖动是如何能够使得通带到阻带的过渡更陡峭，也因此，让我们能够结合普通 Chebyshev 滤波器通带特性与逆 Chebyshev 滤波器的阻带特性。

如果对椭圆型滤波器的传输函数做与从 Chebyshev 到逆 Chebyshev 的转换的类似操作，能够推断出一些重要结论。首先将 Chebyshev 有理函数显式地表示为两个多项式的比：

$$|H(\mathrm{j}\omega)|^2=\frac{1}{1+\varepsilon^2F_{\mathrm{n}}^2(\omega/\omega_{\mathrm{p}})}=\frac{1}{1+\varepsilon^2\dfrac{N(\omega/\omega_{\mathrm{p}})}{D(\omega/\omega_{\mathrm{p}})}}=\frac{D(\omega/\omega_{\mathrm{p}})}{D(\omega/\omega_{\mathrm{p}})+\varepsilon^2N(\omega/\omega_{\mathrm{p}})} \tag{56}$$

其中 N 与 D 分别是 F_{n}^2 的分子与分母多项式。

接下来，用 1 减去功率传输函数，然后对变量 $\omega/\omega_{\mathrm{p}}$ 取倒数：

$$|H_2(\mathrm{j}\omega)|^2=1-|H(\mathrm{j}\omega)|^2=\frac{\varepsilon^2N(\omega/\omega_{\mathrm{p}})}{D(\omega/\omega_{\mathrm{p}})+\varepsilon^2N(\omega/\omega_{\mathrm{p}})} \tag{57}$$

---

[29] Causer（考尔，1900—1945）在 Göttinggen 大学跟随那位极为聪明而又因丢三落四出名的 David Hilbert 开始熟悉了椭圆函数。有一次，Hilbert 突然问一个自己的好朋友，物理学家 James Franck，“你的妻子就像我的妻子那样节俭吗？”Franck 想了一下，反问他，“为什么，你妻子做了什么事不对了？”Hilbert 回答说，“我今天才发现我的妻子不在早餐时给我吃鸡蛋。只有老天才知道这样的事已经发生多久了。”
遗憾的是，关于 Cauer 的故事就没有那么轻松了。他是在第二次世界大战快要结束的那些日子里在苏联占领柏林时被开枪打死的。这让人悲哀地想起阿基米德的结局（见 http://www-ft.ee.tu-berlin.de/geschichte/th_nachr.htm）。

[30] M. E. VanValkenburg, *Analog Filter Design*（模拟滤波器设计），Harcourt Brace Jovanovich, New York, 1982, p. 379。

[31] 严格说来，Chebyshev 多项式是 Chebyshev 有理函数在分母为 1 时的一个特殊情形。

$$|H_3(j\omega)|^2 = \frac{\varepsilon^2 N(\omega_p/\omega)}{D(\omega_p/\omega) + \varepsilon^2 N(\omega_p/\omega)} \tag{58}$$

可以看到 H_3 也是一个与原先 H 同阶的椭圆型滤波器。考虑到椭圆型滤波器在通带与阻带是最小化极大值最优的，而最优的唯一性意味着 H 与 H_3 有相同的极点与零点。通过比较分子，推论出 N 的根与 D 的根互成倒数。因此，计算 D 的根（这些是整个滤波器的零点所在）等于计算 N 的根。只要把这个知识与指定的通带抖动值结合，就有了所有来完成传输函数的全部信息。

就如与椭圆型滤波器有关的所有其他事一样，对于所需阶数的公式的推导是十分困难的，而且会使我们陷入过多繁复的数学推导中（读者可能已经这样认为了）。如果愿意抛开对公式从何而来的好奇性，依然可以进行所需要的计算：

$$n = \frac{F(m)F(1-m')}{F(m')F(1-m)} \tag{59}$$

其中 $F(m)$ 是完整的第一类椭圆积分（其定义与计算的数值方法请参考 22.7 节），m 是一个归一化了的阻带频率的函数，

$$m = 1/\omega_S^2 \tag{60}$$

而 m' 为阻带与通带抖动参数的函数

$$m' = (\varepsilon_s/\varepsilon_p)^2 \tag{61}$$

尽管对于 $F(m)$ 的一个严格解析表达式并不存在，一个十分精确的近似是[32]

$$F(m) \approx (a_0 + a_1 m + a_2 m^2 + a_3 m^3 + a_4 m^4) + (b_0 + b_1 m + b_2 m^2 + b_3 m^3 + b_4 m^4)\ln(1/m) \tag{62}$$

其中不同的系数都在表 22.8 中列出。对于 $0<m<1$，近似的误差不大于 2×10^{-8}。注意参数 m 简单的就是 22.7 节中椭圆参数 k 的平方：

$$m = k^2 \tag{63}$$

表 22.8　对完全第一类椭圆积分近似的系数

a_n	b_n
$a_0=1.386\ 294\ 361\ 12$	$b_0=0.5$
$a_1=0.096\ 663\ 442\ 59$	$b_1=0.124\ 985\ 935\ 97$
$a_2=0.035\ 900\ 923\ 83$	$b_2=0.068\ 802\ 485\ 76$
$a_3=0.037\ 425\ 637\ 13$	$b_3=0.033\ 283\ 553\ 46$
$a_4=0.014\ 511\ 962\ 12$	$b_4=0.004\ 417\ 870\ 12$

一旦需要的阶数 n 确定，下一个目标就是找到归一化零点的位置。这里同样的，只要不关心公式的推导，计算过程没有那么难。存在着很多准确度很高、用于椭圆函数计算的数值方法的近似，它们能直接给出零点的位置（而且它们只是 ω_s/ω_p 值的函数）。形式上，这些零点的位置为

$$\pm j\frac{1}{k\left\{\mathrm{cd}\left[(2i-1)\dfrac{K}{n},k\right]\right\}},\quad i=1,2,\cdots,\mathrm{Int}\left(\frac{n}{2}\right) \tag{64}$$

---

[32] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*（数学函数手册），Dover，New York，1965。

同样地，存在着很多方法来从该表达式中得到实际的数值解，一个简单的(同时又是十分精确的)求零点位置的解析公式为

$$\pm \mathrm{j}\frac{1}{\sqrt{k}}\frac{F_4(i)}{F_3(i)},\quad i=1,2,\ldots,\mathrm{Int}\left(\frac{n}{2}\right) \tag{65}$$

函数 F_3 和 F_4 为

$$F_3(i)=F_{34e}(i)-F_{34o}(i) \tag{66}$$

$$F_4(i)=F_{34e}(i)+F_{34o}(i) \tag{67}$$

其中

$$F_{34e}(i)\approx \alpha^i+\alpha^{(8n-3i)}+\alpha^{(8n+5i)}+\alpha^{(32n-7i)}+\alpha^{(32n+9i)} \tag{68}$$

$$F_{34o}(i)\approx \alpha^{(2n-i)}+\alpha^{(2n+3i)}+\alpha^{(18n-5i)}+\alpha^{(18n+7i)}+\alpha^{(50n-9i)}+\alpha^{(50n+11i)} \tag{69}$$

而

$$\alpha^{2n}=\lambda+2\lambda^5+15\lambda^9+150\lambda^{13}+1707\lambda^{17}+\ldots \tag{70}$$

其中

$$\lambda=\frac{1}{2}\left[\frac{1-\sqrt{k}}{1+\sqrt{k}}\right] \tag{71}$$

前面这一系列公式十分接近地(但并不一致地)沿用了 Orchard 所使用的符号习俗[33]。这些近似具有的项的数目超出了在进行任何实际滤波器设计时所需要的项的数目，但比如说作为用于数据表格时，依然是十分简单的。

零点位置的确定使得之后的实际网络的综合变得可行，只是通过采用过去几十年来发展起来的许多方法的之一[34]。

与逆 Chebyshev 滤波器的情形相同，可以选择如何在沿着滤波器的路径分配共振陷阱的频率。在综合过程中，不奇怪会发现一些选择会带来困难，比如要求有负电感器。经验表明将频率最低的零点设置到处于滤波器中心位置的共振器，然后逐个交替地向外决定结构，可以在很大程度下减轻(但并不是消除)这种困难。例如，如果有五个零点频率，以频率增长作序，将它们从 1～5 编号，然后最佳的实现方式很可能是将谐振器的调谐次序以 4-2-1-3-5 的方法来综合。LADDER 这个程序用这样的综合方法来克服实现中的问题。

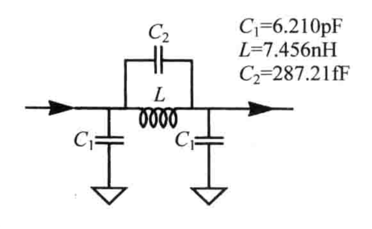


图 22.24　低通 1GHz 椭圆型滤波器例子

继续那个低通的例子，LADDER 综合结果在图 22.24 中显示。对这个滤波器运行 Monte Carlo 仿真，结果如

---

[33] H. J. Orchard, "Computation of Elliptic Functions of Rational Fractions of Quarterperiod"(四分之一周期的有理函数的椭圆函数的计算。译者注：quarter period 是一个特殊函数，在椭圆函数理论中用到)，*IRE Trans. Circuit Theory*, December 1958, pp. 352-355。尽管在嗣后发表的文章中，Orchard 宣布放弃采用解析式的数值计算技术(基于 θ 函数)，而提倡采用迭代的算法(这可不是一件小事，因为这些方法具有显式地揭示 Chebyshev 与椭圆滤波器的极点间的关系的优点)，但依然去采用解析式会更方便些，比如构建可查的数据表。

[34] 一个特别巧妙的方法来自 Pierre Amstutz "Algorithms for Elliptic Filter Design on Small Computers"(可在小型计算机上运行的椭圆型滤波器设计的算法)，*IEEE Trans. Circuits and Systems*, December 1978。遗憾的是，Amstutz 选择了一个椭圆函数的特别的归一化方法。而且他使用的符号习惯对初入门者来理解他文章中的详细数学过程太困难。他通过提供他的滤波器设计算法完整的源程序(用 Fortran 语言)来补偿这个缺陷。依然，他的算法有时会产生不可实现的设计。

图 22.25 所示。我们看到，椭圆型滤波器在我们讨论过的所有的实现方案中具有到阻带的最陡峭的过渡。在这个特定的例子中，阻带响应的最大回跳为一个最大值 -46dB(在图上并不可见)。

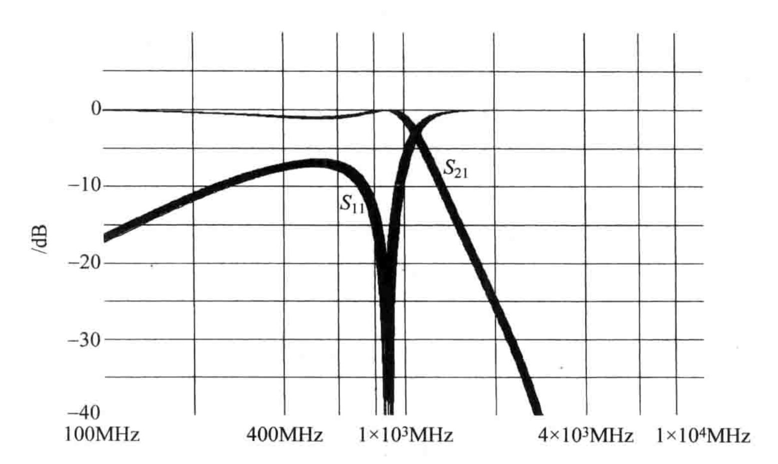


图 22.25　图 22.24 所示的椭圆型滤波器的蒙特卡罗仿真(抖动方差容差为 5%)

22.5.5　Bessel-Thomson 最平坦延迟滤波器

到目前为止，我们研究的滤波器是从满足幅度频响的角度来设计的。由于这些设计方法因此忽略了相位的行为，许多这样的滤波器可能在它们的通带上表现出严重的延时。也就是说，滤波器对输入信号的所有傅里叶分量的延时量并不一定相同。当这些彼此错位的傅里叶分量组合成输出信号时，波形就会产生严重的失真。这种色散行为在数字系统中是一个特别着重关注的问题，在那里为了避免码间干扰，脉冲形状需要被保持良好。

由于常数的时延意味着相移与频率呈线性关系，在一个相位响应最接近线性近似的滤波器中，色散最小。正如传统滤波器的设计忽略相位行为而专注于幅度的响应，我们在设计最大化平坦时延滤波器的相位响应时，将忽略幅度响应的特性。

这种滤波器传输函数的最为巧妙与有效的推导当归功于 Storch[35]。首先我们注意到一个单位时延的传输函数简单地就是

$$H(s)=\mathrm{e}^{-S} \tag{72}$$

然后又注意到 e^{S} 可以分解为如下两个双曲函数的和：

$$\mathrm{e}^{S}=\sinh s+\cosh s \tag{73}$$

下一步将 $\sinh s$ 与 $\cosh s$ 分别表示为简单的级数和，将两者相除可以得到 $\coth s$ 的一个连续分数展开：

$$\sinh s=s+\frac{s^{3}}{3!}+\frac{s^{5}}{5!}+\ldots \tag{74}$$

以及

㉟ L. Storch, "Synthesis of Constant Time Delay Ladder Networks Using Bessel Polynomials"(采用贝塞尔多项式的常数时延梯状网络的综合), *Proc. IRE*, v. 42, 1954, pp. 1666-1675。

$$\cosh s = 1 + \frac{s^2}{2!} + \frac{s^4}{4!} + \ldots \tag{75}$$

因此

$$\coth s = \frac{1}{s} + \cfrac{1}{\cfrac{3}{s} + \cfrac{1}{\cfrac{5}{s} + \cfrac{1}{\cfrac{7}{s} + \ldots}}} \tag{76}$$

连续的分数展开因为可以容易地产生逐级改善的近似而十分吸引人。只需要有足够的项来达到所需要的滤波器阶数时，简单地截断阶数就可以了。因此将分子与分母多项式相加就提供了对 $\sinh s$ 与 $\cosh s$ 和的近似，而根据公式(73)这个和也即 e^s。Storch 的了不起的发现(包括相关证明)是这样形成的近似具有最平坦的时延响应，而那些对 e^s 直接进行传统级数展开得到的结果不能取得这个结果。同样重要的是，他确认了这些多项式属于贝塞尔(Bessel)先前研究过的一类多项式。尽管所有关于这个令人惊叹的工作，对应的这类滤波器类型还是被称为 Bessel, Thomson 或 Bessel-Thomson，因为 W. E. Thomson 大约五年之前就发表了一篇关于这类滤波器的论文，虽然没有如此漂亮的方法也没有意识到与 Bessel 多项式的关系[38]。最后这一点是非常重要的，因为 Bessel 多项式可由常见的(尽管是对数学家而言)递推公式得到。对于与优化了延时问题有关的这一特定类的 Bessel 多项式而言，可以使用下面简单的递推关系来避免计算继续分数的必要：

$$P_n(s) = (2n-1)P_{n-1}(s) + s^2 P_{n-2}(s) \tag{77}$$

其中前两个多项式为

$$P_1(s) = s + 1 \tag{78}$$

与

$$P_2(s) = s^2 + 3s + 3 \tag{79}$$

可以作为递推起始所需要的值。

前六个 Bessel 多项式的系数如表 22.9 所示(我们忽略了首项系数，因为对于所有的情况它都是 1)。

表 22.9 贝塞尔多项式系数

n	a_5	a_4	a_3	a_2	a_1	a_0
1						1
2					3	3
3				6	15	15
4			10	45	105	105
5		15	105	420	945	945
6	21	210	1260	4725	10 395	10 395

一个细节是通过这两种方法的任一种得到的多项式通常没有单位的直流值，因此必须提供一个适当的归一化常数，因此

[38] W. E. Thompson, "Delay Networks Having Maximally Flat Frequency Characteristics"(具有最大化平坦频率特性的延迟网络), *Proc. IRE*, pt. 3, v. 96, 1949, pp. 487-490。

$$H(s)=\frac{P_n(0)}{P_n(s)} \tag{80}$$

另一个细节是需要对频率进行归一化：由多项式迭代公式产生的不同的 $H(s)$ 源自于近似一个单位时延，因而带宽并不被控制到一个统一的值。为了将带宽归一化到一个常数值，对每个 $H(s)$ 计算 −3dB 拐角频率，而后将 s 进行相应的缩放来产生 1-rps 拐点频率。这些归一化的前提目标是为了得到一个具有特定的拐点频率的滤波器，同时提供一个对统一时延的较好的近似。如果更加关心的是提供在一个尽可能宽的带宽内对一个时延的最好近似，就不需要进行这样的归一化[37]。

注意上面描述的这些方法直接产生了传输函数而不是响应幅度的平方，这样就减少了花在寻找在 22.6 节描述的网络综合方法中的 Hurwitz 多项式的工作负荷。遗憾的是，不像 Butterworth 与 Chebyshev 的情况，看起来不存在着能够直接给出元件参数的对于 Bessel 滤波器的简单公式。因此依赖于用标准的综合配方来产生归一化元件参数表中的数据值（表 22.10，更多的数据可以在 22.8 节中找到）。

表 22.10　1-rps Bessel-Thomson 低通滤波器的归一化元件值

$C_1(L_1)$	$L_2(C_2)$	$C_3(L_3)$	$L_4(C_4)$	$C_5(L_5)$	$L_6(C_6)$
2.0000					
2.1478	0.5755				
2.2034	0.9705	0.3374			
2.2404	1.0815	0.6725	0.2334		
2.2582	1.1110	0.8040	0.5072	0.1743	
2.2645	1.1126	0.8538	0.6392	0.4002	0.1365

看起来同样没有简单的设计公式来选择，基于比如说对预定标称延时的可以接受的偏差的阶数。然而，对于不同阶的 Bessel 滤波器延时-频率曲线的观察揭示出延迟带宽 $\mathrm{BW_D}$（当延时相比低频时值的下降 3dB 时）与标称延时的乘积可以由一个简单（但非常粗略）的近似给出：

$$(\mathrm{BW_D})(\mathrm{T_D})\approx n-0.4 \tag{81}$$

一个对可以允许的延时偏移的更紧的指标要求滤波器的阶数增加，公式(81)至少提供了某个指导。一般而言，当延时带宽增加到比正常的带宽大时，非线性的相位逐渐变得没有那么重要，因为只要那些不恰当的傅里叶分量幅度已大大地衰减了，其确切的延迟量已不再是那么重要了。不论哪一种情况，滤波器设计软件，例如 LADDER，能轻松产生 Bessel 滤波器的设计。

让我们用一个例子来结束对各种设计的讨论。任意地选择一个五阶、1GHz 滤波器，其两个端口都使用了并联的电容。我们将表 22.10 中对 50Ω 系统的数值进行去归一化，这个频率给出的元件值如表 22.11 所示。

---

[37] 然而，假若所希望的只是一个时延近似，一些好得多的替代方式是存在的。Bessel-Thompson 滤波器是一个最小相位滤波器（所有的零点都在左半平面），因而在近似的能力上是受限的。提供允许使用右半平面的零点，可以将对一个给定的滤波器的阶数加倍。被称为 Padé 近似的一类近似方法通过提供成镜像对的极点与零点来利用这个观察结果。所对应的网络在本质上是全通的，因此不提供滤波作用。

表 22.11 五阶 3dB,1GHz Bessel-Thomson 低通滤波器去归一化后的元件值

C_1	L_2	C_3	L_4	C_5
7.1881pF	8.8411nH	2.5592pF	4.0362nH	0.5548pF

这个滤波器的仿真得到的结果是－1dB 点发生在 587MHz,因此我们需要进行第二次重新归一化。在这个例子中,需要将所有元件数值乘以 0.587 将－1dB 频率提升到 1GHz,见表 22.12。

表 22.12 五阶 1dB,1GHz Bessel-Thomson 低通滤波器元件的最终值

C_1	L_2	C_3	L_4	C_5
4.2194pF	5.1897nH	1.5022pF	2.3692nH	0.3256pF

对该滤波器的 Monte Carlo 仿真(图 22.26)揭示了对参数变动的相当低的敏感度。然而这个滤波器直到在 5GHz 左右才提供 30dB 的衰减,而设计的目标是 3GHz。为了满足阻带的指标,需要增加阶数。

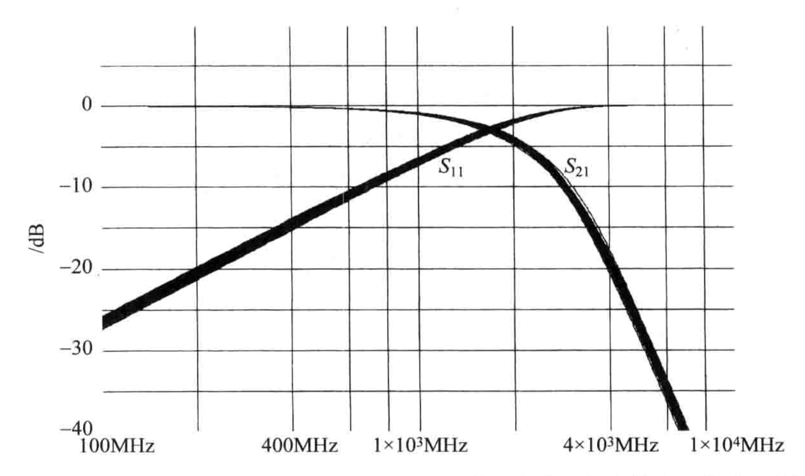


图 22.26 五阶 1dB,1GHz Bessel-Thomson 低通滤波器的蒙特卡罗仿真(5%)

在比较不同的滤波器类型时,我们发现了一个普遍的趋势:滤波器幅度频响越理想,时延的平坦度就越差。例如,Bessel 滤波器在幅度平坦度方面相对较差,但是时延均匀性就很好。满足类似的幅度特性的椭圆型滤波器通常会有更差(有时差得很多)的时延行为。一个直观的原因是椭圆型滤波器用了较少的极点来满足幅度频响的指标要求。较低阶因此意味着较少的总的潜在相移。而一个较小的总的可能实现的相移又意味着在一个给定的频率范围内减弱了提供线性增加的相移的能力。

篇幅的限制使得我们仅能简单提及其他近似线性相位行为的可能方法。Bessel-Thomson 滤波器提供了最大化平坦时延,因此在延时的角度上对应着 Butterworth 的幅度响应滤波器。可以将这个延时与幅度的这种对偶的想法推广出去,来想像一下比如 Chebyshev 滤波器对应的延时对偶,这种滤波器提供了一个对常数延时的等抖动近似。这样的滤波器对于给定的阶数能够提供更大的延时带宽。

另一个选择是对滤波器差的相位响应用另外一个被称为相位均衡器来进行补偿。这样一个均衡器对于所有的频率都有常数的增益(因此会是一个全通网络)，这就如同一段导线，但是它具有一个可控的非零相移(这就与导线不同了)。这样的滤波器在右半平面有零点，而在左半平面有镜像极点。在右半平面上的零点与左半平面的零点有相同的幅度响应，但是相位的行为则与左半平面的极点相同。因此，这些都不是最小相位的网络。

作为最后一个注解，重要的是去注意到产生基于低通原型的其他滤波器形状的各种变换仅仅保留了幅度频的特性，而并不具备对相位行为明显的控制。因此，低通带通变换遗憾地不保留 Bessel-Thomson 滤波器的恒时延特点。需要有一个对线性相位带通滤波器的直接综合来得到最大化平坦的带通延迟特性。

在下一章中，将研究如何将这些集总滤波器转化为可由微带线元件实现的形式。

22.6 附录 A：网络综合

从滤波器频率响应的指标规定到一个可实现的网络的过程或许并不像我们所希望的那样直截了当。事实上，一般说来，存在着很多可能的路径，每一个路径有其自己独特的在直观性、计算复杂度和鲁棒之间的折中选择。不幸的是，我们不能指望对如此复杂的一个课题提供完整的阐述(这在本章里，一直是持续的遗憾)，不过我们能在这里对整个综合过程提供这个简单的大纲，使这个过程不再那么神秘，同时帮助一些感兴趣的读者去参考相关的文献。作为一个辅助的介绍，我们也提供一些关于椭圆函数的背景知识，因为这个题目在大部分电气工程专业的课程体系中被格外地忽略掉了。

现代网络综合的诞生经常被追溯到在 1931 年发表 Otto Brune 的博士论文[38]。Sidney Darlington(Darlington 对的发明者，他还有很多其他成就)将 Brune 对单端口的工作扩展到了双端口网络的综合，在 1939 年发表了后来常被引用的博士论文[39]。来自 Foster、Cauer 和其他一些学者的更多的重要贡献又帮助这个工作牢牢地确定了它的地位。慢慢地，研究的重点逐渐从可实现性的证明转向了发展实际的综合方法。随着这些方法不断得到发展，它们能够计入非理想的元件的效应，以及兼顾对滤波器响应更为复杂的限制的能力也不断得以提高。由于这些进步，加上机器计算容易获得，现代滤波器综合终于在 20 世纪 50 年代后期开始占主导地位。

我们现在考虑上述工作中很小很小的一部分，讨论能综合仅由纯无损元件组成的滤波器的一种可能方法。

无损阶梯网络综合的一个步骤

一个经典的综合无损阶梯网络的方法依赖于一个无损二端口网络的传输函数和它的输入阻抗。这种方法是合理的，因为综合具有特定阻抗的单端口网络的那些方法是首先被提出的。因此，首先确定根据所期望的滤波器转移函数(而这又是由所选的该类滤波器的设计

[38] "Synthesis of a Finite Two-Terminal Network whose Driving Point Impedance Is a Prescribed Function of Frequency"(驱动端阻抗为给定频率函数的有限两端网络的综合)，*J. Math. Phys.*, v. 10, 1931, pp. 191-236。

[39] "Synthesis of Reactance 4-Poles which Produce Prescribed Insertion Loss Characteristics"(能产生预先设定的插入损耗特性的电抗 4-极点的综合)，*J. Math. Phys.*, v. 18, September 1939, pp. 257-353。

限制所决定的)所隐指的输入阻抗函数,而后将输入阻抗函数按照连续分数级数的形式展开。在许多实际的相关例子中,元件的数值可以容易地从观察展开式中的各项中读出(这种便捷正是促使 Cauer 对连续分数级数的热情推崇的原因)。幸运的是,Butterworth、Chebyshev 和 Bessel-Thomson 滤波器都能属于这类可以用这种方法来进行相对较为简单的综合过程的滤波器,前提是假设我们使用理想、无损的电感和电容。然而,使用有损元件以及/后者带有虚零点(如逆 Chebyshev 和 Cauer 类型)的滤波器综合,很遗憾地需要使用比我们这里所描述的更复杂的步骤[40]。

继续进行讨论,首先用连续分数级数的形式推导一个阶梯网络的输入阻抗的表达式,也借此解释一下 Cauer 对此所显示的热情。特别是,考虑图 22.27 中的阶梯网络,这个网络是用电阻 R 作为终端负载(这里一个 LC 阶梯是作为一个特定的例子,但稍后的推导能被更广泛地应用到)[41]。

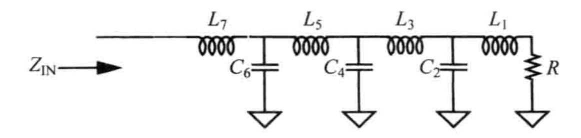


图 22.27　终端接有负载的 LC 阶梯状网络

对于这样一个网络,其输入阻抗可以通过几个方式来得到。对于大多数人来说,自然的选择是从输入端开始,然后朝着负载端来进行,但是从负载开始逆向到输入端是有其优点的。因此从 C_2 右侧看到的阻抗开始:

$$Z_1 = sL_1 + R \tag{82}$$

同时,Z_1 与 C_2 阻抗的并联组合具有导纳

$$Y_2 = sC_2 + \frac{1}{Z_1} = sC_2 + \frac{1}{sL_1 + R} \tag{83}$$

再向后再退一步得到从 C_4 右侧看上去的阻抗,有

$$Z_3 = sL_3 + \frac{1}{Y_2} = sL_3 + \cfrac{1}{sC_2 + \cfrac{1}{sL_1 + R}} \tag{84}$$

明显地,可以继续向后推进直到最终得到输入阻抗的表达式。通过对表达式形式的观察,读者可以看到为什么连续分数级数这个词是一个合适的描述。更重要的是,在级数展开中的每一步的第一项都神奇地给出了单个梯形网格元件的值。如果能够将输入阻抗用这样的形式表示,就在实质上能通过目测容易地读出各元件的值。

为了在网络综合中运用这个观察结果,需要找到一种方式去将输入阻抗显式地与将要被综合的传输函数联系在一起。幸运的是,在概念上这是比较直截了当可以做到的。功率传输函数告诉我们输入功率中多少会被传输到负载上,然后可以轻松地计算出多少功率是

[40] 如果需要了解有关更一般的网络综合方法的细节,参看 M. E. VanValkenburg, *Introduction to Modern Network Synthesis*(现代网络综合导论), Wiley, New York, 1960。也可以参考在 A. Zverev, *Handbook of Filter Synthesis*(滤波器综合手册), Wiley, New York, 1967 中提供的丰富的设计数据。

[41] 这里列出的基本方法当然可以被一般化以兼容不相等的源与负载端的终断(对偶数阶的 Chebyshev 滤波器的综合特别有用),但是为了简单化起见,我们将只考虑相等的电阻的情况。

被反射回至源。再利用与史密斯图中输入阻抗与反射系数关系的相同的方法可以推出滤波器输入阻抗。

实现这套操作的 Darlington 方法是简单地将能量守常数律应用到如图 22.28 所示系统开始的。由于阶梯网络本身是无损的，传输到负载的平均功率必须等于提供给网络输入端的平均功率。认识到功率仅仅传输给输入阻抗 Z_{in} 的实部 R_{in}，那么功率平衡的标准隐指

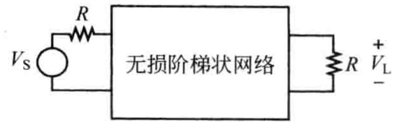


图 22.28　终端接有负载的无损阶梯状网络

$$|I_{in}|^2 R_{in} = \frac{|V_L|^2}{R} \tag{85}$$

同时注意到输入电流简单地为

$$I_{in} = \frac{V_S}{R + Z_{in}} \tag{86}$$

这可以使我们写出

$$\left(\left|\frac{V_S}{R+Z_{in}}\right|^2 R_{in} = \frac{|V_L|^2}{R}\right) \Rightarrow \left|\frac{V_L}{V_S}\right|^2 = \frac{RR_{in}}{|R+Z_{in}|^2} \tag{87}$$

在已经将希望得到的滤波器传输函数以已知的终端负载电阻与网络的输入阻抗形式表示，下面间接地通过一个中间过程（这是由 Darlington 最先给出）来确定网络阻抗，这是利用了如下的一个辅助函数[42]：

$$|A(j\omega)|^2 = 1 - 4\left|\frac{V_L}{V_S}\right|^2 = 1 - 4\frac{RR_{in}}{|R+Z_{in}|^2} = \frac{|R+Z_{in}|^2 - 4RR_{in}}{|R+Z_{in}|^2} \tag{88}$$

我们接着对最终公式的分子部分展开，将输入阻抗依虚部与实部用显式表示出来，而得到下面一系列的公式：

$$\frac{|R+Z_{in}|^2 - 4RR_{in}}{|R+Z_{in}|^2} = \frac{(R+R_{in})^2 + X_{in}^2 - 4RR_{in}}{|R+Z_{in}|^2} = \frac{R^2 + R_{in}^2 + X_{in}^2 - 2RR_{in}}{|R+Z_{in}|^2} \tag{89}$$

这又可以简化为

$$|A(j\omega)|^2 = \frac{|R-Z_{in}|^2}{|R+Z_{in}|^2} \tag{90}$$

令 $s=j\omega$ 并求解 Z_{in}，依靠这个辅助函数与终端电阻，最终得到输入阻抗的两个表达式：

$$Z_{in} = R\frac{[1-A(s)]}{[1+A(s)]} \tag{91}$$

$$Z_{in} = R\frac{[1+A(s)]}{[1-A(s)]} \tag{92}$$

注意这两个表达式对应两个独立但等价的网络。进一步观察到，如果终端电阻为单位值，则 Z_{in} 的两个表达式实际上是互为倒数的。因此在整个综合过程中都使用归一化的数值能简化对这两个解中元件值的计算。在综合过程中最费时的部分完成之后，只需要进行逆归一化来作为最后一步。

---

[42] 这个辅助函数初看起来可能有点神秘，但从其形式可以注意到这其实恰好就是作为频率函数的反射功率。就如在正文中所阐明的，在反射与负载阻抗（这里就是带有负载的梯形网络的输入阻抗）间存在着一个一一对应的关系，因此知道了反射系数等价于知道了梯形网络的输入阻抗。读者可以回想起来这个反射系数与阻抗的一一对应关系就是史密斯圆图的基础。

一个示范性的例子可以帮助阐述方法中的细节,同时揭示出某些需要小心处理的地方。特别是,考虑 22.5.1 节中的 Butterworth 滤波器例子。所要求滤波器的归一化的幅度响应平方为

$$|H(j\omega)|^2 = \frac{1}{1+\varepsilon^2 (\omega/\omega_p)^{2n}} \tag{93}$$

为了保持与这个附录中推导的一致性,注意到

$$4\left|\frac{V_L}{V_S}\right|^2 = |H(j\omega)|^2 = \frac{1}{1+\varepsilon^2 (\omega/\omega_p)^{2n}} \tag{94}$$

其中系数 4 是计入在直流时 $|V_L/V_S|$ 值为 0.5 这样一个事实,这是源于我们这个滤波器的模块图。计算辅助函数是一个十分直接的过程,即在归一化到一个单位通带频率之后,得到

$$|A(j\omega)|^2 = 1-4\left|\frac{V_L}{V_S}\right|^2 = 1-|H(j\omega)|^2 = \frac{\varepsilon^2\omega^{2n}}{1+\varepsilon^2\omega^{2n}} \tag{95}$$

可以看到,当滤波器传输函数 $|H(j\omega)|$,归一化到单位直流增益之后,Darlington 辅助函数定义中的 4 被消掉了,使得数学的预算得以简化一些。

现在令 $s=j\omega$(因此 $\omega^2=-s^2$),同时认识到这样的替换使得 $|A(j\omega)|^2=A(s)A(-s)$。亦即,这个辅助函数模的平方为两个项的乘积,每一项的极点与另一项对应的极点在代数上互为逆(即符号值不同)。这里遇到第一个细微之处:我们是在寻求函数 $A(s)$,但是只是给定了其模的平方值。我们实际上只需要 $|A(j\omega)|^2$ 的那一部分,其在右平面上没有极点或零点。用网络理论家的语言来说就是,我们要求一个 Hurwitz(最小相位的)多项式。因此需要找到公式(95)中分子与分母多项式的根,然后扔掉在右半平面的根,再依据剩下的根重新构建一个多项式来得到所要的 $A(s)$[43]。如果多项式不能用目测或其他简便的方式来进行因式分解,那就需要使用数值方法来求根。这一步可能存在的问题是,求根的数值方法的精确度在某些时候可以比实际需要的元件值精确度还要高很多[44]。在公式(95)中分子不需要特别的处理,因此求根的问题在这里是限于对分母的。

在这个特别的例子中,我们很幸运没有太多的困难就可以求得根。从下式开始:

$$|A(j\omega)|^2 = \frac{\varepsilon^2\omega^{2n}}{1+\varepsilon^2\omega^{2n}} \Rightarrow A(s)A(-s) = \frac{\varepsilon^2 s^{2n}}{1+\varepsilon^2 s^{2n}} \tag{96}$$

从中可以看到通过求解下式中的 s 就可以得到希望的极点

$$1+\varepsilon^2 s^{2n} = 0 \tag{97}$$

这等价于求解

$$s^{2n} = -\frac{1}{\varepsilon^2} \tag{98}$$

因此寻找能够给出纯负值的 s。这个观察对于求解很重要,因为将数 -1 表示成 e 的一个带

[43] 假若,如这个例子中发生的,在坐标原点处存在着多个根,就保留其中的一半(它们总是成对发生的)。

[44] 在众多的可以使用的求根算法中,滤波器的设计者绝大多数都倾向于使用 Laguerre(拉盖尔)方法。尽管缺乏有关这个算法的一般性收敛性质的任何正式的证明,大量的经验已经证明这个颇受推崇的方法在实践中是极其鲁棒的。特别是它能很好地处理有重复根的十分具有挑战性的问题。要了解这个算法的详细描述,可参看 W. H. Press 等 *Numerical Recipes*(数值分析方法大全),Cambridge University Press(任意一个版本)。另一本有价值的参考书是 H. J. Orchard, "The Laguerre Method for Finding the Zeros of Polynominals"(求多项式根的拉盖尔方法),*IEEE Trans. Circuits and Systems*, v. 36, no. 11, November 1989, pp. 1377-1381.

恰当虚指数的指数项可以有

$$s^{2n} = -\frac{1}{\varepsilon^2} = \frac{1}{\varepsilon^2} e^{j(2k+1)\pi} \tag{99}$$

其中参数 k 的取值范围从 $0\sim 2n-1$（实际上，任意 $2n$ 个连续整数都能达到同样的效果）。

该方程的解因而为

$$s_k = \frac{1}{\varepsilon^{(1/n)}} e^{[j(2k+1)\pi]/(2n)} \tag{100}$$

对于这个四阶的例子（$n=4$）

$$s_0 = \frac{1}{\varepsilon^{(1/n)}} e^{j(\pi/8)} \tag{101}$$

$$s_1 = \frac{1}{\varepsilon^{(1/n)}} e^{j(3\pi/8)} \tag{102}$$

这第一二个根是在右半平面上，因而被忽略。

接下来，得到

$$\left(s_2 = \frac{1}{\varepsilon^{(1/n)}} e^{j(5\pi/8)} \right)$$

$$\Rightarrow \tau_2 = \frac{1}{s_2} = \varepsilon^{(1/n)} e^{j(-5\pi/8)} \approx \varepsilon^{(1/n)} [-0.38268 - j0.92388] \tag{103}$$

$$\tau_3 = \varepsilon^{(1/n)} e^{j(-7\pi/8)} \approx \varepsilon^{(1/n)} [-0.92388 - j0.38268] \tag{104}$$

$$\tau_4 = \varepsilon^{(1/n)} e^{j(-9\pi/8)} \approx \varepsilon^{(1/n)} [-0.92388 + j0.38268] \tag{105}$$

$$\tau_5 = \varepsilon^{(1/n)} e^{j(-11\pi/8)} \approx \varepsilon^{(1/n)} [-0.38268 + j0.92388] \tag{106}$$

这就确定了四个在左半平面上的根。

注意选择用时间常数来表示根。这个变换能够通过将具有下列形式的项相乘来重新构建辅助函数

$$(\tau s + 1) \tag{107}$$

容易使得采用正确的系数表示 $A(s)$ 来让其在高频时具有所要求的单位幅度值（见公式(96)）。

丢弃剩下的两个根，

$$s_6 = \frac{1}{\varepsilon^{(1/n)}} e^{j(13\pi/8)} \tag{108}$$

$$s_7 = \frac{1}{\varepsilon^{(1/n)}} e^{j(15\pi/8)} \tag{109}$$

因为它们的实部为正数。

现在有四个相关的根，可以完成辅助函数的推导：

$$A(s) = \frac{\varepsilon s^4}{(\tau_2 s - 1)(\tau_3 s - 1)(\tau_4 s - 1)(\tau_5 s - 1)} \tag{110}$$

将负载电阻归一化到单位电阻得到期望的输入阻抗函数：

$$Z_{in} = \frac{1 - \dfrac{\varepsilon s^4}{(\tau_2 s - 1)(\tau_3 s - 1)(\tau_4 s - 1)(\tau_5 s - 1)}}{1 + \dfrac{\varepsilon s^4}{(\tau_2 s - 1)(\tau_3 s - 1)(\tau_4 s - 1)(\tau_5 s - 1)}} \tag{111}$$

$$Z_{\text{in}}=\frac{1+\dfrac{\varepsilon s^4}{(\tau_2 s-1)(\tau_3 s-1)(\tau_4 s-1)(\tau_5 s-1)}}{1-\dfrac{\varepsilon s^4}{(\tau_2 s-1)(\tau_3 s-1)(\tau_4 s-1)(\tau_5 s-1)}} \tag{112}$$

这些表达式可以进一步简化为

$$Z_{\text{in}}=\frac{(\tau_2 s-1)(\tau_3 s-1)(\tau_4 s-1)(\tau_5 s-1)-\varepsilon s^4}{(\tau_2 s-1)(\tau_3 s-1)(\tau_4 s-1)(\tau_5 s-1)+\varepsilon s^4} \tag{113}$$

$$Z_{\text{in}}=\frac{(\tau_2 s-1)(\tau_3 s-1)(\tau_4 s-1)(\tau_5 s-1)+\varepsilon s^4}{(\tau_2 s-1)(\tau_3 s-1)(\tau_4 s-1)(\tau_5 s-1)-\varepsilon s^4} \tag{114}$$

为了从这里开始减小推导的工作量，我们将忽略公式(114)所指的综合选择。接下来的讨论会清晰地表明如何在这个公式的基础上建立网络，因此忽略掉第二个选择没有概念上的损失。

将合适的根代入公式(113)，再经过简化得到

$$Z_{\text{in}}=\frac{[k^2s^2+s(2ka)+1][k^2s^2+s(2kb)+1]-\varepsilon s^4}{[k^2s^2+s(2ka)+1][k^2s^2+s(2kb)+1]+\varepsilon s^4} \tag{115}$$

其中

$$k=\varepsilon^{(1/4)};\quad a\approx 0.382\,68;\quad b\approx 0.923\,88 \tag{116}$$

进一步简化得到

$$Z_{\text{in}}=\frac{\varepsilon s^4+s^3[2k^3(a+b)]+s^2[2k^2(1+2ab)]+s[2k(a+b)]+1-\varepsilon s^4}{\varepsilon s^4+s^3[2k^3(a+b)]+s^2[2k^2(1+2ab)]+s[2k(a+b)]+1+\varepsilon s^4} \tag{117}$$

最终有

$$Z_{\text{in}}=\frac{s^3[2k^3(a+b)]+s^2[2k^2(1+2ab)]+s[2k(a+b)]+1}{2\varepsilon s^4+s^3[2k^3(a+b)]+s^2[2k^2(1+2ab)]+s[2k(a+b)]+1} \tag{118}$$

下一步要产生 Z_{in} 的一个连续分数表达式。在综合除法的每一步，对高阶的多项式用低阶的多项式来除，在整个过程中，如果需要的话将被运算量求逆(你会注意到随着除法的进行，项是成对地被丢弃的)。例如，因为公式(118)的分子的阶数较低，作为展开的第一步，我们用它来除分母的多项式。

进行综合的除法产生了以下的一系列项，每一项交替地产生一个电感性的电抗与电容性的导纳(选择哪一个作为开始并不重要)：

$$s\frac{k}{a+b}=sL_1 \tag{119}$$

$$s\frac{k(a+b)}{2ab}=sC_2 \tag{120}$$

$$s\frac{8k(ab)^2}{(a+b)(4ab-1)}=sL_3 \tag{121}$$

以及最后是

$$s\frac{k(a+b)(4ab-1)}{2ab}=sC_4 \tag{122}$$

通过对元件值的逆归一化来完成设计，我们得到在 Butterworth 滤波器这一节给出的“L 为先的”版本的滤波器，见表 22.13。由于互易性与相同的输入输出端口的终断，我们可以互换输入输出的作用，而不改变滤波器的传输函数。这样做，会产生“C 为先”的滤波器版本。这种灵活度有时是很有价值的，因为它能够吸收在两个端口处存在的不同(或者是大小或者

是特性上)的寄生参数。

表 22.13　四阶 1dB,1GHz Butterworth 低通滤波器($Z=50\Omega$)元件参数

L first
$L_1=5.1441\text{nH}$
$C_2=4.9675\text{pF}$
$L_3=12.419\text{nH}$
$C_4=2.0576\text{pF}$

从这个例子中可以看到手工的综合设计要求相当的工作量,也有很多发生错误的可能(想象一下如果设计的是十一阶的滤波器,你就会体会到了)。但是,这个过程的直接性很容易实现对自动化。存在着大量的滤波器设计软件可以帮助完成这个困难的工作。特别具有吸引力的是 LADDER 这个程序,这一方面由于它是免费的,同时也是因为它以教学为导向的特点,包括随设计的进行将综合每一步的结果呈现的选项[45]。

22.7　附录 B　椭圆积分、函数与滤波器

如本章正文所述的那样,Cauer 滤波器在数学上是基于椭圆函数的。但是如果说椭圆这个词与 Chebyshev 滤波器中极点沿着(半)椭圆的分布无关,这个词从何而来？如何这个词与滤波器设计相关？这个附录给出了一个简短的有关椭圆函数的介绍,其目的是在某些基本的滤波器手册中的知识与更高等的滤波器处理之间建立一个桥梁。对数学细节不感兴趣的读者可以(甚至是鼓励)跳过这些内容。

22.7.1　为什么它们是"椭圆"

要解决一个物理问题的需求往往激起了数学上的某一个特定的课题,但是通常研究的方向之后就远离了初衷。有些数学家经常鼓励这样的过程,他们不希望他们的工作与物理现实有太过明显的联系。对于椭圆函数与积分同样如此。

如果读者上过微积分的课,下面这个问题(和相应的求解方法)会是比较熟悉的：对于一个半径为 a 的园,计算一段圆弧的长度,例如从 $15°\sim38°$。基于微积分的方法是对微分圆弧长度的方程在指定的区间内进行积分。这并不是一件困难的事,经常是被用作家庭作业的练习或作为在入门的微积分课程课堂上演示的例子(这部分是由于结果可以用一个基于非微积分的简单方法的推导来加以验证)。

然而,如果是尝试来回答求一个椭圆的弧长或者周长的问题,情况就从很简单变为了几乎不可能。你得到一个积分,采用标准微积分课程中教的方法是无法求值的。事实上,已经被严格证明,这些椭圆积分的通用求值需要用数值方法来解决。在没有计算工具可以被普遍使用前,大量的表(或许是用级数展开的方法计算得到)是大部分工程师们能得到实际结果(而不是符号形式)的唯一方式。

---

[45] R. D. Koller and B. Wilamowski, "LADDER—A Microcomputer Tool for Passive Filter Design and Simulation"(LADDER——一个用于无源滤波器设计与模拟的微计算机工具), *IEEE Trans. Education*, v. 39, no. 4, November 1996, pp. 478-487.

计算上的困难并不是对椭圆函数理解的唯一障碍。例如，*Numerical Recipes*（数值分析方法大全）[46]的作者，含蓄地指出，“对使用椭圆积分更难的事情之一是在文献中围绕着这个课题的符号用法的如此之繁多。”存在很多不同的符号标记。而且，如今叫做椭圆积分的量曾经被称作椭圆函数。如我们将要看到的，后者现在指用椭圆积分定义的一类函数。在阅读这方面的文献时，应注意这种称谓上的区别。

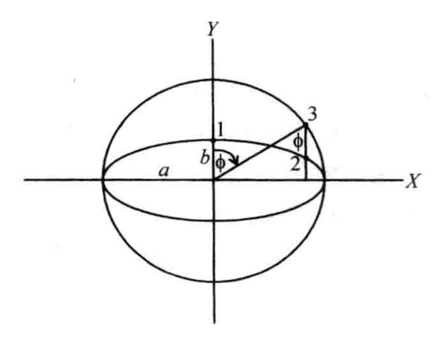


图 22.29　内切于一个圆的椭圆

这里，采用 E. Jahnke 与 F. Emde 使用的符号[47]。通过参考图 22.29 可以了解这种选择的原因，在图中将椭圆内切在一个半径为 a 的圆内。此外，我们将原点与圆上的点 3 相连，将点 3 垂直投影到 x 轴上。最后，注意用顺时针方向定义正的角度，相对于纵轴。

椭圆的方程为

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1 \tag{123}$$

其中 a 与 b 分别为半长轴与半短轴的长度。

在椭圆上的弧长，比如从点 1 到点 2，可以首先通过计算点 2 的坐标来容易得到。依据目测，x 的坐标就是

$$x=a\sin\phi \tag{124}$$

要找到点 2 的 y 坐标则需要稍微多一点的努力。将公式(124)代入公式(123)，求解 y 得到

$$y=\frac{b}{a}\sqrt{a^2-x^2}=\frac{b}{a}\sqrt{a^2-(a\sin\phi)^2}=b\cos\phi \tag{125}$$

或许通过几何方式有更直接的方式来得到这个结果，但现在我们已经满意了能用随便哪个方法得到了解。

现在，弧长可以一般性地表示为

$$s=\int\sqrt{\mathrm{d}x^2+\mathrm{d}y^2} \tag{126}$$

对于积分中的变量用恰当的量取代，得到

$$s=\int\sqrt{(a\cos\phi)^2+(b\sin\phi)^2}\,\mathrm{d}\phi=\int\sqrt{a^2[1-(\sin\phi)^2]+(b\sin\phi)^2}\,\mathrm{d}\phi \tag{127}$$

经过简化并将符号整理得更清晰一些(用 ψ 表示积分变量同时保留 ϕ 作为积分上界)，得到

$$s=a\int_0^{\phi}\sqrt{1-k^2\ (\sin\psi)^2}\,\mathrm{d}\psi \tag{128}$$

其中 k 为椭圆的离心率：

$$k=\sqrt{1-\frac{b^2}{a^2}}=\frac{\sqrt{a^2-b^2}}{a} \tag{129}$$

---

[46] W. H. Press *et al*. (Cambridge University Press, 1986)

[47] 对于椭圆积分的一个出色的导论可以参考 H. W. Reddick 与 F. H. Miller 的 *Advanced Mathematics for Engineers*(工程师用高等数学)，3rd ed.，Wiley，New York，1955。这本教科书采用了与 Jahnke and Emde 的 *Tables of Functions with Formulae and Curves*(用公式与曲线表示的函数表)(Dover，New York，1945)一样的符号规范。

从图中显而易见，对于圆的特殊情况 k 为 0，且不会超过 1[48]。当半长轴 a 为单位长度，这个积分被称为第二类椭圆积分：

$$E(k,\phi)=\int_0^{\phi}\sqrt{1-k^2\ (\sin\psi)^2}\,d\psi \tag{130}$$

E 的选择可以理解为在普通几何的意义下，直接用来表示椭圆的弧长，这可以帮助记忆。

因为前面的推导都比较清晰，因此在物理上也是相关的，现在必须开始一些带有技巧的使问题模糊化的做法。在椭圆函数这个范畴里，离心率 k 被重新命名为模数，而上限 ϕ 则称为幅度。将这个已经是含混的课题变得更加含混，在很多文本与表格中采用了模角(就是模 k 的反正弦值)而不是 k 本身：

$$k=\sin\theta \tag{131}$$

除了模数之外，我们还经常用到互补模数：

$$k'=\cos\theta \tag{132}$$

这样

$$k'=\sqrt{1-k^2} \tag{133}$$

将这里考虑的积分划为第二类仅仅是由于数学家(主要是来自 Abel 与 Jacobi，在 Legendre 的先驱性工作之后)在归类上所做的顺序选择。

现在回到椭圆，很明显将幅度设置为 $\pi/2$，计算积分得到的是在一个象限中的圆弧长度。由于椭圆的对称性，知道这个弧长之后就可以得到整个周长。对应的积分便称作完全的第二类椭圆积分。被记为 $E(k)$，或简单地就是 E：

$$E(k)=\int_0^{\pi/2}\sqrt{1-k^2\ (\sin\psi)^2}\,d\psi \tag{134}$$

整个椭圆的周长于是为 $4aE(k)$。

第一类椭圆积分是类似的，不过积分函数为第二类积分函数的倒数：

$$F(k,\phi)=\int_0^{\phi}\frac{1}{\sqrt{1-k^2\ (\sin\psi)^2}}\,d\psi \tag{135}$$

符号 F 的用法提示我们所对应的积分为第一类。这些积分是我们在设计椭圆滤波器中遇到的积分，也是在严格求解一个摆动的摆锤的振荡周期的积分[49]。

22.7.2 椭圆函数

早些时候，我们提到过椭圆积分曾经被称为椭圆函数，而后面这个叫法现在是用来专用椭圆积分来定义的一类函数。现在定义这些椭圆函数，回想起相角 ϕ 也被称为第一类椭圆积分的幅度。

如果积分被认为是这个角度的函数，那么其逆函数即为幅度。根据传统的做法，在这种情况下的积分通常不用 F 而是用 u 来表示(同样，不要问为什么)：

$$\phi=\mathrm{am}\ u \tag{136}$$

---

[48] 不过，这一支数学所做的扩展可以允许超过单位(即 1)的 k 值(以及复数值)，但在那种情况下，要保持与问题的原本起源的简单联系是十分困难的。

[49] 因为读者无疑急切地想知道，其实在真空中的一个摆锤的振荡周期严格地由 $4(L/g)^{1/2}[F(k,\pi/2)]$ 给出，其中 L 是摆锤的长度，g 是重力加速度，k 是摆锤相对于垂直轴的最大角偏移值一半的正弦值。

如果想显式地提醒读者 u 也取决于模数 k，也可以写作

$$\phi = \mathrm{am}(u, \mathrm{mod}k) \tag{137}$$

因为幅度本身就是相角，它可以被用作标准三角函数的自变量，这样最后定义三个椭圆函数：

$$\sin\phi \stackrel{\mathrm{def}}{=} \sin \mathrm{am}\ u \stackrel{\mathrm{def}}{=} \mathrm{sn}\ u \tag{138}$$

$$\cos\phi \stackrel{\mathrm{def}}{=} \sqrt{1-(\mathrm{sn}\ u)^2} \stackrel{\mathrm{def}}{=} \cos \mathrm{am}\ u \stackrel{\mathrm{def}}{=} \mathrm{cn}\ u \tag{139}$$

$$\sqrt{1-k^2\ (\sin\phi)^2} \stackrel{\mathrm{def}}{=} \Delta\phi \stackrel{\mathrm{def}}{=} \mathrm{dn}\ u \tag{140}$$

当 $k=0$ 时，前两个椭圆函数分别退化为常规的 $\sin\phi$ 与 $\cos\phi$。当 $k=1$，函数(140)蜕变为传统的双曲正切。因此在实际意义上，椭圆函数可以被认为是广义的三角函数，这些函数随着模数从零变到一从普通的连续地变到双曲函数。这个行为的连续性是许多用于精确计算椭圆函数值方法的基础。

这些函数的不同比例又定义了其他的椭圆函数，函数第一个与第二个字母分别来自组成这个函数的分子与分母的第一个字母。例如，

$$\mathrm{cd}\ u = \frac{\mathrm{cn}\ u}{\mathrm{dn}\ u} \tag{141}$$

是一个这样的函数。它正好直接源自计算椭圆滤波器零点的位置。

22.7.3　椭圆函数的数值计算

如前所述，椭圆积分不能用那些基本的函数(例如，代数、三角和指数函数，以及它们的逆函数)来计算。一个直截了当的(但不是很有效的)数值方法，可以将积分函数展开为幂级数(使用，例如二项式理论)，并且将每一项分别积分。对于需要提醒的读者，二项式级数是：

$$(1+x)^n = 1+nx+\frac{n(n-1)x^2}{2!}+\frac{n(n-1)(n-2)x^3}{3!}+\dots \tag{142}$$

这正是最初建立椭圆积分表时所采用的类似方法。在计算能力很强的今天，用算法实现而不是利用查表经常是更为容易。

提供完全椭圆函数的级数展开是最为容易的，这也是我们在这里给出的全部。尽管一些中间的步骤没有在这里给出，如果读者决定要做，应该是能够填补这些中间过程的。至今为止总结的推导步骤最终导致下述相对简约的表达式，如下面的级数所示：

$$F(k)=\frac{\pi}{2}\left[1+\left(\frac{1}{2}\right)^2k^2+\left(\frac{1\cdot 3}{2\cdot 4}\right)^2k^4+\left(\frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\right)^2k^6+\dots\right] \tag{143}$$

$$E(k)=\frac{\pi}{2}\left[1-\left(\frac{1}{2}\right)^2k^2-\left(\frac{1\cdot 3}{2\cdot 4}\right)^2\frac{k^4}{3}-\left(\frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\right)^2\frac{k^6}{5}-\dots\right] \tag{144}$$

我们稍后将给出计算效率要高得多的方法，但是直接的级数展开对于不那么接近 1 的 k 值收敛速度相当快，在这种情况下使用这个级数展开是完全可行的。另一个极好的替代方法是采用在本章正文中给出的近似级数——如果 20ppb 最差情况误差是可以接受的(这几乎总是如此)。

对于大小不是 $\pi/2$ 的情况，按照我们简单的解决方法通常需要重复使用下面的迭代公式来得到$\sin^n\psi$ 的积分：

$$\int(\sin\psi)^n\mathrm{d}\psi=-\frac{(\sin\psi)^{n-1}\cos\psi}{n}+\frac{n-1}{n}\int(\sin\psi)^{n-2}\mathrm{d}\psi \tag{145}$$

用这种方式计算不完整椭圆积分因此可能是个多少麻烦的事。更糟的是，得到的级数收敛速度会相当慢(尤其是对于模数接近1的情况)。因此，就不奇怪有一些聪明的学者经过艰苦工作来得到了更为有效、收敛更快的一些方法。然而，解释它们的推导过程比说明它们的操作过程要复杂得多，因此我们简单地给出下列算法，而对读者指出能发现更详细细节的参考文献[50]。

为了计算不完整的第一类积分，(F_n, k_n, ϕ_n)，首先定义如下的一个由三个数组成的序列：

$$F_{n+1} = \frac{F_n}{1 + \sqrt{1 - k_n^2}} \tag{146}$$

$$k_{n+1} = \frac{2}{1 + \sqrt{1 - k_n^2}} - 1 \tag{147}$$

$$\tan(\phi_{n+1} - \phi_n) = (\tan\phi_n)(\sqrt{1 - k_n^2}) \tag{148}$$

其中初值设置为 $\phi_0 = \phi, k_0 = k$ 以及 $F_0 = 1$。

如果如下表示，公式(148)可能看起来更直观些(但是在计算上效率却较低)

$$\phi_{n+1} = \phi_n + \arctan\left[(\tan\phi_n)\sqrt{1 - k_n^2}\right] \tag{149}$$

其中“atan”表示反正切。

进行迭代直到 $k_n = 0$ 在某个容差(如10^{-10})成立。

$$F(k, \phi) = F_N \phi_N \tag{150}$$

其中 N 为结束迭代时 n 的值。收敛的速度通常是如此之快，要得到十分高的准确度通常只需要在十次以内的迭代。

对于第二类不完整的椭圆积分也有类似的计算步骤，但是因为在滤波器设计中不需要这类积分，这里忽略对它们的考虑。

计算完整积分(包括两种类型)也很容易用快速收敛的方法进行。这里首先定义一对数(a_n, g_n)，然后相继计算它们的算术和几何平均值：

$$a_{n+1} = \frac{a_n + g_n}{2} \tag{151}$$

$$g_{n+1} = \sqrt{a_n g_n} \tag{152}$$

初值为 $a_{-1} = (1 + k)$ 和 $g_{-1} = (1 - k)$。

进行迭代，同时跟踪期间得到的这对数值。会发现算术和几何均值会收敛到一起。继续迭代直到在某个容差(如10^{-10})内有 $a_n = g_n$。那么，

$$F(k) = \frac{\pi}{2a_N} \tag{153}$$

其中 N 为最终迭代结束时的 n 值，同时

$$E(k) = \frac{F(k)\left[2 - (a_0^2 - g_0^2) - 2(a_1^2 - g_1^2) - 4(a_2^2 - g_2^2) - \cdots\right]}{2} \tag{154}$$

其中不同平方差项的系数简单的就是2的幂。

最后，任何一个想从电机工程师的角度来理解椭圆函数的读者绝对需要阅读 Harry Orchard 写的有关这个议题的极为出色的一系列论文。其中最近的一篇文章总结了半个世

[50] 在这里给出的算法是由在 J. A. Ball 的 *Algorithms for RPN Calculators*(用于 RPN 计算器的算法)(Wiley, New York, 1978)那些算法改动而来。而 Ball 则是实现了一个由 Landen 最早提出的最有用的方法。

纪以来他在椭圆滤波器设计方面积累的知识与了解，这对初学者格外有用[51]。

22.7.4 有用的近似公式

尽管本章主要是围绕滤波器设计的，我们通过考虑计算椭圆弧长的问题引入了椭圆函数。现在对已经给出的较为正式的这些方法提供一些补充来作为结束。下面给出一个简单的对椭圆周长(circumference)的闭式近似，这是由富有天赋的 *Ramanujan* 得到的几个公式之一：

$$\text{circumference} \approx \pi\left[3(a+b)-\sqrt{(3a+b)(3b+a)}\right] \tag{155}$$

他的公式可以用来推出

$$E(k) \approx \frac{\pi}{4}\left[3(1+\sqrt{1-k^2})-\sqrt{(3+\sqrt{1-k^2})(3\sqrt{1-k^2}+1)}\right] \tag{156}$$

这个公式给出的值对于从零到稍微大于 0.93 的 k 值与发表的表中给出的值在小数点后四位相符。误差随 k 接近 1 增加，但仍然小于 0.5%。

22.8 附录 C　通用低通滤波器的设计表

这里对几种在本章正文中讨论到的低通滤波器的元件参数以表格形式给出。直到第九阶的常数 k，Butterworth，Chebyshev 和 Bessel 低通滤波器的归一化了的元件值都在表中列出。在所有下面这些表格中，归一化是按照 1Ω 和 1rps 通带频率(或者对于常数 k 和 m 导出式型而言，是截止频率)进行的。对于不同的阻抗级别，要将元件的阻抗相应地按比例缩放。例如，对于一个 50Ω 系统，要将所有的电感乘以 50，而对所有的电容除以 50。对于不同的截止频率，用与频率成反比的方式对每个元件的阻抗进行缩放。例如，对于截止频率为 10Grps，将每个电感和电容除以10^{10}。

如本章正文所述，可以从低通的那些原型推导出高通、带通和带阻滤波器。相关的变换在本附录末给出。值得强调的是这些变换只保留了幅度频响的形状，但是不一定能保持相位的响应。注意到这一点对 Bessel-Thomson 滤波器格外有关。不幸的是，使用标准的变换方法无法从低通原型推出线性的相位响应的带通的滤波器。要得到线性相位响应的带通滤波器，必须采用其他方法，例如直接综合。

所有的这些表是通用的，这是指可以选择或者从一个并联电容或串联电感开始。后面这个选择(从电感开始)在括号里表示。也可以自由地交换任何这些滤波器的输入和输出端口，而对传输特性没有改变。这个结论即便对非对称的滤波器(例如偶数阶的 Butterworths)也是成立的，因为这个特性并不是基于对称结构，而是基于网络的互易特性，同时假设相等的源及负载终断。

在图 22.30 中，L 与 C 的值是对应 1rps 的原型情况，这也解释了在图中第一列给出的初始频率归一化。元件值也都是归一化到 1Ω，因此需要最后的反归一化来得到实际的阻抗值以完成设计：

$$L = RL_n \tag{157}$$

[51] H. J. Orchard and A. N. Willson, "Elliptic Functions for Filter Design"(用于滤波器设计的椭圆函数), *IEEE Trans. Circuits and Systems I*, v. 44. No. 4, April 1997, pp. 273-287.

$$C = (1/R)C_n \tag{158}$$

表 22.14 常数 k 低通滤波器的归一化元件值

$C_1(L_1)$	$L_2(C_2)$	$C_3(L_3)$	$L_4(C_4)$	$C_5(L_5)$	$L_6(C_6)$	$C_7(L_7)$	$L_8(C_8)$	$C_9(L_9)$
1.000	2.000	1.000						
1.000	2.000	2.000	2.000	1.000				
1.000	2.000	2.000	2.000	2.000	2.000	1.000		
1.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	1.000

表 22.15 Butterworth 低通滤波器($\varepsilon=1$)的归一化元件值

$C_1(L_1)$	$L_2(C_2)$	$C_3(L_3)$	$L_4(C_4)$	$C_5(L_5)$	$L_6(C_6)$	$C_7(L_7)$	$L_8(C_8)$	$C_9(L_9)$
2.000								
1.414	1.414							
1.000	2.000	1.000						
0.7654	1.848	1.848	0.7654					
0.6180	1.618	2.000	1.618	0.6180				
0.518	1.414	1.932	1.932	1.414	0.518			
0.445	1.247	1.802	2.000	1.802	1.247	0.445		
0.390	1.111	1.663	1.962	1.962	1.663	1.111	0.390	
0.347	1.000	1.532	1.879	2.000	1.879	1.532	1.000	0.347

表 22.16 抖动为 0.1dB 的 Chebyshev 低通滤波器的归一化元件值

$C_1(L_1)$	$L_2(C_2)$	$C_3(L_3)$	$L_4(C_4)$	$C_5(L_5)$	$L_6(C_6)$	$C_7(L_7)$	$L_8(C_8)$	$C_9(L_9)$
1.032	1.147	1.032						
1.147	1.371	1.975	1.371	1.147				
1.181	1.423	2.097	1.573	2.097	1.423	1.181		
1.196	1.443	2.135	1.617	2.205	1.617	2.135	1.443	1.196

表 22.17 抖动为 0.5dB 的 Chebyshev 低通滤波器的归一化元件值

$C_1(L_1)$	$L_2(C_2)$	$C_3(L_3)$	$L_4(C_4)$	$C_5(L_5)$	$L_6(C_6)$	$C_7(L_7)$	$L_8(C_8)$	$C_9(L_9)$
1.596	1.097	1.596						
1.706	1.230	2.541	1.230	1.706				
1.737	1.258	2.638	1.344	2.638	1.258	1.737		
1.750	1.269	2.668	1.367	2.724	1.367	2.668	1.269	1.750

表 22.18 抖动为 1.0dB 的低通滤波器的归一化元件值

$C_1(L_1)$	$L_2(C_2)$	$C_3(L_3)$	$L_4(C_4)$	$C_5(L_5)$	$L_6(C_6)$	$C_7(L_7)$	$L_8(C_8)$	$C_9(L_9)$
2.024	0.994	2.024						
2.135	1.091	3.000	1.091	2.135				
2.167	1.112	3.094	1.174	3.094	1.112	2.167		
2.180	1.119	3.121	1.190	3.175	1.190	3.121	1.119	2.180

表 22.19　1rps Bessel-Thomson 低通滤波器的归一化元件值

$C_1(L_1)$	$L_2(C_2)$	$C_3(L_3)$	$L_4(C_4)$	$C_5(L_5)$	$L_6(C_6)$	$C_7(L_7)$	$L_8(C_8)$	$C_9(L_9)$
2.000								
2.1478	0.5755							
2.2034	0.9705	0.3374						
2.2404	1.0815	0.6725	0.2334					
2.2582	1.1110	0.8040	0.5072	0.1743				
2.2645	1.1126	0.8538	0.6392	0.4002	0.1365			
2.2659	1.1052	0.8690	0.7020	0.5249	0.3259	0.1106		
2.2656	1.0956	0.8695	0.7303	0.5936	0.4409	0.2719	0.0919	
2.2649	1.0863	0.8639	0.7407	0.6306	0.5108	0.3770	0.2313	0.0780

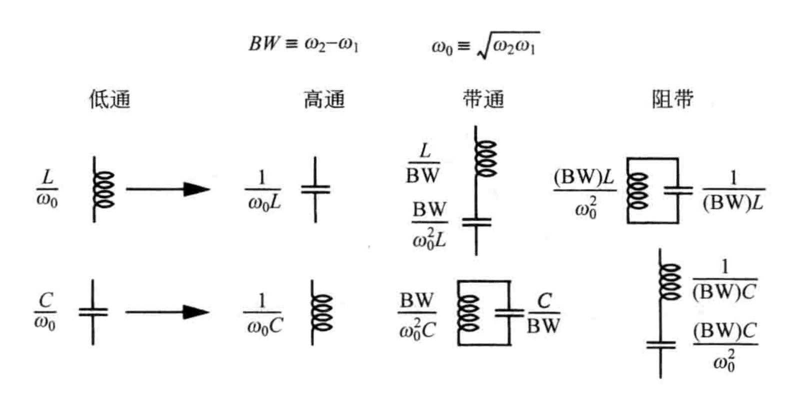


图 22.30　从低通原型到其他形状变换的总结

第23章 微带线滤波器

23.1 背景

在这一章里,我们考虑设计分布式滤波器的一系列方法。所有这些方法都是从集总元件的滤波器原型中得出来的,以便尽可能地利用现有的有关集总元件滤波器设计的大量文献[①]。在一些例子中,这类向分布式形式的变换只包括了用传输线近似直截了当地替代集总元件。在另一些例子中,变换是基于这样的认识,即原先与集总元件电路相联系而发展起来的某些原理有更广泛的应用。总体而言,这些转换在通带特性上一般能提供十分接近于集总实现的类似的特性。然而,分布式滤波器带来了一个特别的挑战:在高于阻带频率的一些频点,频响会弹升起来,尽管集总设计的响应是单调衰减的。传输线周期性的阻抗特性不可避免地带来了这种凹形的频响[②]。来确定与仔细地调整这些模式成为了很多微波滤波器工程师的某些首要考虑的事。经验告诉我们,即便经过大量的迭代,我们依然需要接受这些不可避免的现实:这样得到的滤波器与它们的集总原型相比总是会不一样的。在将期待值适当降低之后,现在可以来进行一些推导了。

23.2 从集总原型得到的分布式滤波器

用于得到一些最早期的微波滤波器的一个直接的方法对于很多应用依然工作得很好:只要简单地将集总原型中的分立电感器和电容器用等效的用传输线段做成的元件替代就可以了[③]。就如在第7章中所讨论的,当传输线段在长度上只是电波长的一个合适的小部分时,传输线作为集总元件的近似是非常好多。一段末端开路的传输线可以很好地被用作一个电容,而一小段末端短路的传输线,其特性是一个电感。

然而,一个重要的考虑是,随着频率增加,所有传输线最终都不再相对于波长是非常短的了。与之发生的阻抗变动随之改变了滤波器的响应。例如,一个微带低通滤波器频响可能会在通常的阻带内回升。由于这种失真的响应并不是低通滤波器所独有的,我们需要认

---

① 存在着不依赖于对集总元件原型作近似的直接设计方法。这些直接方法可以得到更好的滤波器结果,但是是以采用直观上不容易明白或解释的步骤为代价。我们这里集中在非直接方法,因为除了要求最为苛刻的应用外,这些方法能足够好的工作,并且能用直观上能说清楚的概念来容易地加以解释。

② 然而,不应该得出这样的结论,即总的频响因此是周期性的。只有当这些元件都具有同样的长度,才会有简单的周期性。

③ 见,比如 G. L. Ragan(编辑),Microwave Transmission Circuits(微波传输线)(MIT Rad. Lab. Ser., v. 9), chap. 10(by Fano and Lawson), McGraw-Hill, New York, 1948。同时可参考由 Seymour Cohn 在 *Very High Frequency Techniques*(极高频率技术),Radio Research Laboratory, McGraw-Hill, New York, 1947, 第2卷中写的一章。这些早期的实现不可避免地会不完整,但是从历史上来说是重要的。

真地判断任何一个提出的实现方式，来保证所有不希望的响应在幅度或频点位置都不是有害的。

为了推导一个将集总原型转换到分布式滤波器的简单方法，首先回想一下对一小段末端开路的传输线的输入阻抗近似为

$$Z \approx \frac{Z_0}{j\omega(l/v)} \tag{1}$$

因此等效电容为

$$C = \frac{l}{vZ_0} = \frac{l\sqrt{\varepsilon_{r,eff}}}{cZ_0} \tag{2}$$

在 FR4 上的 50Ω 传输线大致给出 1.3pF/cm，但是我们同样关心能够得到的最大的实用电容值是多大。采用具有相对低的阻抗的传输线，实践上可以得到大约为 4pF/cm 的值。在 1GHz 频率下，将元件的尺度限制在远小于波长的要求致使长度必须比 15cm 小不少，这相对于总电容大约在 5～10pF 之下。这个电容的极限值随着频率的增加呈平方关系减小，因为电容与面积成正比，而面积则与波长的平方成正比。

类似地，对于一根末端短路的短传输线的电感，有

$$L = \frac{lZ_0}{v} = \frac{lZ_0\sqrt{\varepsilon_{r,eff}}}{c} \tag{3}$$

如我们经常引用到的，对于 FR4 上最窄(或最高阻抗)的实际可做到的传输线，电感通常的取值大致在 1nH/mm 的量级。同样，在 1GHz，我们发现最大的可实现的电感值为大约 10～20nH。这个近似的电感限制与频率成反比。

一个关键的观察是，这些关系即使对于不是以理想开路或短路作为终端的传输线段也是很好的近似。前面的这些公式依旧是相当准确的，只要线段是以近似开路或短路的阻抗终断的，这个阻抗是相对于传输线的特征阻抗而言的：

$$\frac{Z(z)}{Z_0} = \frac{\dfrac{Z_L}{Z_0} + j\tan\beta z}{1 + j\dfrac{Z_L}{Z_0}\tan\beta z} \tag{4}$$

其中采用了坐标 z 是正值的做法④，负载位于 $z=0$ 的位置，同时，

$$\beta = \omega/v \tag{5}$$

因此得到结论，只要 Z_L 与 Z_0 的差异是大的，阻抗便收敛于一些简单的形式。对于 $Z_L \ll Z_0$，长度为 l 的一根短线将给出以下的归一化阻抗：

$$\frac{Z(z)}{Z_0} \approx j\tan\beta l \tag{6}$$

这是具有电感性的。对于 $Z_L \gg Z_0$，

$$\frac{Z(z)}{Z_0} \approx \frac{1}{j\tan\beta l} \tag{7}$$

这为容性。因此为了使近似成立，需要选择尽可能低(或实际能做到的)的 Z_0 来实现一个电

④ 抱歉，我们在书的中间改换了坐标的符号习惯。幸好，只要我们不是在推导过程的中间改换符号，这并没有影响。

容器，而选择尽可能高的 Z_0 来实现一个电感器。

当然不能要求任意高的特性阻抗，因为实际可以制造的传输线宽度是有一个底限的。假设实际制造的误差为 2mil(50μm)并且假设这个误差代表最多是总宽度的 20%，可以假设最小的可制造的宽度为 10mil(250μm)[⑤]。因此，在 1.6mm 厚的 FR4 上，实际的传输线阻抗很少超过 200Ω，而 150Ω 为一个比较常见的最大值。

同时，对线的宽度同样存在着一个最大值的限制，这是因为一个微带线元件的所有各方向的尺寸相对所有目标频率内的波长都需要是小的，以此来保证对集总元件特性较好的近似。相对应隐含的阻抗的下限依赖于工作频率范围，但是作为一个通用的规则，小于 10Ω 的特征阻抗很少被用到，而 15Ω 则为一个比较常见的数值。因此在实现微带滤波器时，要记住在 FR4 基底上，实际可实现的阻抗值通常是在 50Ω 的 3～4 倍之内。

23.2.1 步进阻抗滤波器

或许将分立原型转换成微带线形式的最简单的办法是仅仅使用可以方便(或重复)制造的最窄和最宽的传输线两种。如前面讨论的，窄的传输线是用来实现串联电感，而宽的传输线实现并联电容。在步进阻抗滤波器中，长度调整到必要的程度来产生所要的元件值。因此在这类滤波器中只有三种宽度的传输线(第三种是用于阻抗为 Z_0 的输入/输出传输线)。

一如可能期望的那样，这种转换本身的近似本质限制了它的可用性。阶梯阻抗滤波器也因此最好用于那些应用，那里相比于集总滤波器原型频响，较大的误差能够被容忍，或者是你不在乎不断地迭代来改进滤波器精度。这些误差的严重性通常当截止频率被超过而变大，因为实际的传输线段的分布本征特性随频率的增加而更为明显。采用精心的设计，步进阻抗和集总参数滤波器在低于截止频率和其附近的频率范围内特性非常相近。然而超过截止频率，阶梯阻抗滤波器通常不如原型衰减的那么快，而且的确最终的阻带衰减或许不能满足设计要求。同时，滤波器的频响可能会出现大量的失真的通带。由于单个传输线段通常是不等长的，滤波器频响不会显示出简单的周期性。

在说明了以上这些后，让我们看看如何用阶梯阻抗结构来构建一个低通滤波器。作为一个特别的例子，假设希望截止频率为 1GHz，同时使用恒定 k 原型作为微带线滤波器的基础。如果原型有两个完整的 T 形单元(或三个完整的 π 形单元)，那么步进阻抗滤波器将有七个单元，分别对应于集总原型中的七个元件。进一步假设可以实现的最大和最小的传输线阻抗分别为 15Ω 与 200Ω[⑥]。为了与集总原型元件值相符，我们要求一个大小为 15.915nH 的电感，我们用最窄的线来实现，其长度由下式给出：

$$l = \frac{vL}{Z_{0,\max}} = \frac{f\lambda L}{Z_{0,\max}} \tag{8}$$

这对应在截止频率处电感的归一化长度大约为 29.647°。[⑦] 类似的，那些主要的 6.3662pF 的电容需要有长度为

⑤ 假若成本不是一个问题，我们可以要求制造厂商在硬的基底上提供窄到 1mil(25μm)的线宽。

⑥ 这样一个高的值当所用的衬底不是 FR4 时可能不容易实现。大多数(幸好不是全部)微波衬底具有高得多的电介质常数，使得实现高阻抗的传输线尤其困难。

⑦ 在截止频率处实现阻抗匹配多少是带有任意性的，但是因为我们经常十分关心在截止频率附近的行为，这又是一个好的选择。同时，我们提供了一些看上去是荒唐的小数点位数，这是为了便于那些想独立地进行比较的读者。

$$l = vCZ_{0,\min} = f\lambda CZ_{0,\min} \tag{9}$$

这对应于大约 28.742°的归一化长度。那些半单元终断匹配电容正好为该长度的一半。

这个滤波器的版图大致如图 23.1 所示,其中集总原型也给出以作为参考。应用 Puff 这个程序对该滤波器的仿真给出在 1GHz 截止频率处的衰减为 3.9dB,及 −3dB 带宽为 989MHz。由于 Puff 没有考虑伴随长度步长改变带来的场失真,其结果会不同于从场求解器得到的结果。特别急剧的步长(即宽度)的改变在这里会造成十分明显的后果(例如那些源自高阻抗单元的等效长度变短,正如我们在第 7 章中提到的 T 单元变短)。同样地,当频率高到足以使电容单元的间隔不那么大(如,几个介质层厚度),相互之间的耦合也会改变响应。尽管这么说,我们所关注的基本特性依然可以为 Puff 仿真所反应出来。

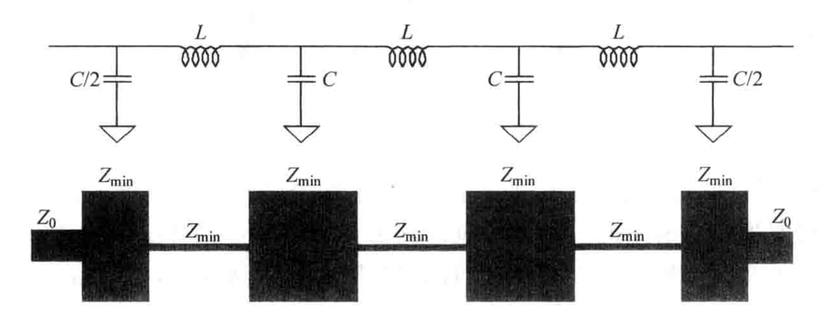


图 23.1　阶跃阻抗滤波器例子(没有严格按照比例画)

仿真(参考图 23.2)显示阶跃阻抗滤波器的性能与恒定 k 原型类似,其在截止频率处的衰减为 10dB 而 −3dB 带宽为 934MHz。最大的阻带衰减超过 60dB,但是没有像集总模型一样呈现出随频率增加的单调性。或许比这些差别更重要是存在着 2～3GHz 宽的失真的通带,对该特定的实现,通带的中心频率在 6.2GHz 附近。当然理想的集总恒 k 滤波器没有这种失真的通带,同时,这样不同的特性在任何分布滤波器、阶跃阻抗或其他方法的实现中都需要加以考虑。

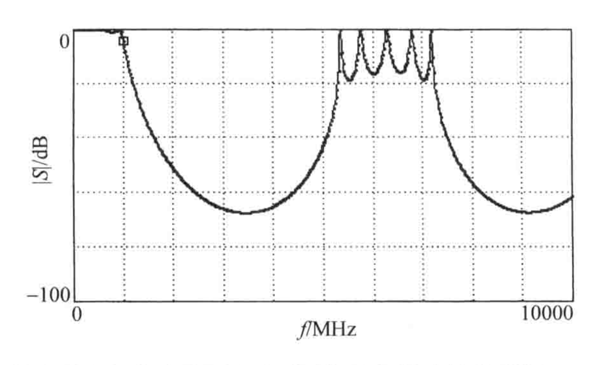


图 23.2　在较宽频率与衰减范围内阶跃阻抗滤波器的特性(源自 Puff 模拟结果)

23.2.2　分支(stub)低通滤波器

一个替代的实现办法继续使用窄传输线作为串联电感,但是用接到该电感型的主干线终端开路的支路来实现电容。如果这些支路可以做得比用在传统的阶梯状阻抗滤波器的传

输线更窄，那么得到的结构会更接近于它所基于的集总原型——至少对于那些频率，在其上支路的长度相对波长依然是短的。

在图 23.3 中，注意到支路被交替地放置在干线的两侧。这种安排不是必须的，但是它的确可以减小不希望的相邻（或甚至距离较远的）线段间的耦合。这种耦合可能以一种不希望看到的方式改变滤波器的频率响应。如之前所建议的那样，达到 4～5 倍的介质宽度的线与线之间的间隔是足以避免这类问题的。

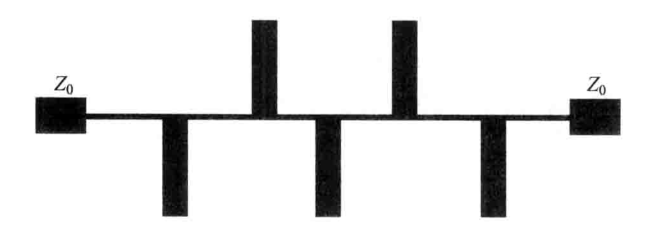


图 23.3　带支路的低通滤波器例子（没有按照比例画出）

这些支路的值（但并一定是它们的形状）是与那些给阶梯状阻抗型滤波器完全相同的，它们仅仅是在安排上有所不同。对这个设计的仿真给出了在 1GHz 处的衰减为 6.5dB，在 965MHz 处的衰减为 3dB。因此，相比较对阶梯状阻抗滤波器变换，这些通带特性多少更接近于集总原型。

离得远一点来看在更宽频率和衰减的动态范围内的滤波器频响（图 23.4），我们看到带支路的滤波器与阶梯状阻抗实现比较在大致相同的频率范围同样有失真的通带。然而那些峰值更小也更窄，这样整体上有更小的由凹形模式产生的发射。再一点是，不同于阶梯状阻抗滤波器，在某个频率区间，阻带的衰减超过 100dB。当然，这样大的衰减在实际中从未发现过，这是因为实际上总是存在的损耗不会产生无穷深的凹陷。即使是传输线可以被假设成没有损耗的，在实际系统中依然不会有十个数量级这样大的隔离。无论如何，设计一个干净的测量来验证这种衰减的存在本身也是一个重要的测量方法和装置的挑战。

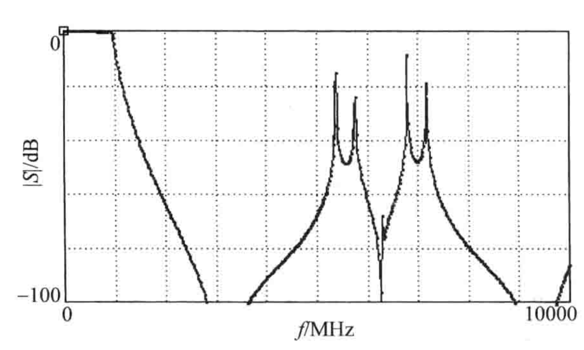


图 23.4　在大的衰减和频率范围内带支路滤波器特性（Puff 模拟的结果）

在这个情况下，我们已经看到带支路的滤波器看上去要比阶梯状阻抗版本更好些。这种优越性是普遍成立的，因为就从集总原型的映射而言，带支路的实现要比阶梯状阻抗更为准确些。然而，这种通用的观察结果并不因此就可以免除进行验证的必要性，尤其是在性能要求高的场合下。

如果所需要的支路要求有不容易实现的低阻抗，那么最好用两个并联的支路来实现，各自置于主线的两侧。然而，需要考虑到支路的总长度，因为会带来麻烦的横向振荡可能在更高的频率下发生，在这些频率时总长度为半波长的奇数倍。

一个经常被问到的问题是在沿着主线的哪一个位置来精确地连接支路。对于电学窄的支路，将连接点的位置视为支路宽度的一半是合理的。在更高的频率或对于非常低阻抗(宽)的支路，如何令人满意地回答该问题就变得困难了。通常的一个解决方法是使用一个扇形的支路，其比较窄的顶端可以大大减小这种不确定性，参考图 23.5。

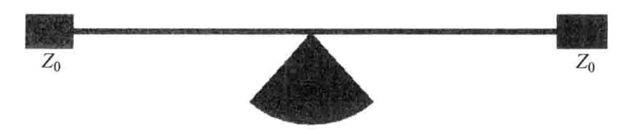


图 23.5　带有扇形支路的滤波器

这时候设计的自由度包括扇形结构的半径、角位移大小和在附着点的宽度(或半径)(太小的宽度会带来过高的串联电阻值)。扇形的支路可以成对使用，成蝴蝶形的排列，以提供非常低的阻抗(数值能容易地做到小于 10Ω)。如图 23.6 中所示，只有不多的支路能沿着主线放置。

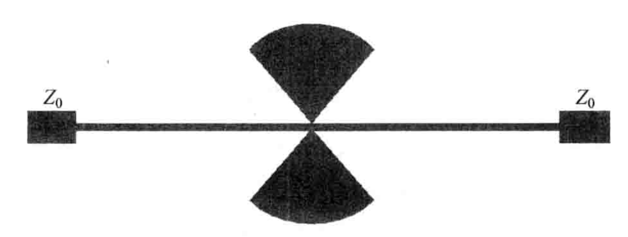


图 23.6　带有蝶形支路的传输线

扇形支路只要其半径相比于波长非常短，它就是一个对电容的很好的近似。其产生一个大的、高质量的电容的能力使得扇形(或蝶形)支路成为电源供电网路旁路电容的十分常见的元件选择，无论是在板级还是在集成电路的实现都是。对那些精度不那么重要的应用，平行板近似通常精度就足够了：

$$C \approx \varepsilon\alpha r_L^2/2H \tag{10}$$

其中假设了外半径比接触点的半径要大很多。公式(10)中的介质常数为全(不是相对的)介质常数，这个近似假设边缘效应可以忽略，因此有效的介电常数和体介电常数本质上是相同的。

对那些十分挑剔的人，可以把弧形支路考虑成其本身是一类特别的传输线的一个部分，就像是一个普通的支路是普通传输线的一部分。很少有文献处理一般化的情况，但是以下的直观的公式可以在 *HP* 过去提供的应用说明中找到[⑧]：

---

⑧ 应用说明(Applications Note)976，“Broadband Microstrip Mixer Design—The Butterfly Mixer”(宽带微带线混频器设计——蝶形混频器)，1980. 在这个说明中，引用了 J. R. Vinding 的“Radial Line Stubs as Elements in Strip Line Circuits”(径向传输线支路作为微带线电路的元件)，IEEE NEREM Record，1967，pp. 108-109。我们在这一章中给出的方程是根据在应用说明中提供的略加修改的版本。

$$X = \frac{HZ_0}{\alpha r_i}\frac{\cos(\theta_i - \psi_L)}{\sin(\psi_i - \psi_L)} \tag{11}$$

其中 X 为支路电抗，H 为介质厚度，α 为扇形支路的角度（以 rad 为单位），r_i 为接触点处的半径，同时

$$Z_0 = \frac{\eta_0}{\sqrt{\varepsilon_r}}\sqrt{\frac{J_0^2(kr_i)+N_0^2(kr_i)}{J_1^2(kr_i)+N_1^2(kr_i)}} \tag{12}$$

$$k = \frac{2\pi\sqrt{\varepsilon_r}}{\lambda_0} \tag{13}$$

$$\theta_i = \arctan\left[\frac{N_0(kr_i)}{J_0(kr_i)}\right] \tag{14}$$

$$\psi_i = \arctan\left[-\frac{J_1(kr_i)}{N_1(kr_i)}\right] \tag{15}$$

$$\psi_L = \arctan\left[-\frac{J_1(kr_L)}{N_1(kr_L)}\right] \tag{16}$$

回想一下 η_0 为真空的阻抗

$$\eta_0 = \sqrt{\mu_0/\varepsilon_0} \tag{17}$$

知道下述这些符号代表的意义或许是有用的：r_L 为总半径，J_n 与 N_n 分别为 n 阶第一类和第二类 Bessel 函数。

23.2.3 椭圆、*m* 导出式和逆 Chebyshev 低通滤波器

如我们在前一章中看到的那样，在频响中具有零点（即有限的传输零点）的集总滤波器（如 m 导出型、逆 Chebyshev 和椭圆滤波器）可以用串联路径上的并联谐振器或用接地的串联谐振器来实现。基于后者结构的微带线滤波器实现起来更方便些。再一次，串联谐振器用小段的窄传输线构造的电感和小段的宽传输线构造的电容来实现，参考图 23.7。

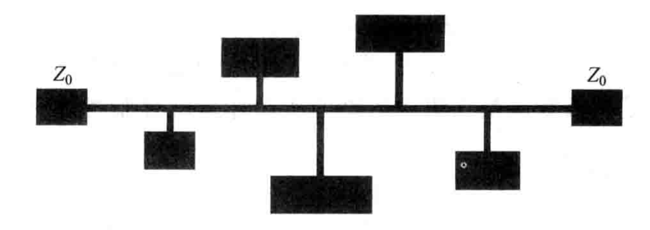


图 23.7　具有有限传输零点的低通滤波器的版图例子（未按比例画）

就如集总椭圆滤波器，最实际的微带线实现是将最低频的谐振器放在中间位置，交替地将每一个接下去的谐振器朝输入/输出端口逐步放置。如果分布式版本是从集总原型推出来的，这个合适的序列应当已经在原型的综合过程中考虑到了。

举一个设计实例来说明，让我们用刚刚总结的方法将一个简单的集总椭圆滤波器转换为分布形式。直接基于 22.5.4 节中给出的例子来实现会有些问题，因为它用一个并联谐振回路作为陷波元件。使用微带线的形式来实现这样一个网络并不是一件容易的事。幸运的是，陷波也可以通过一个串联共振支路并联到滤波器的主干路来实现，而且滤波器设计工具，如 LADDER，也能综合这种选择。综合的结果在图 23.8 中给出。

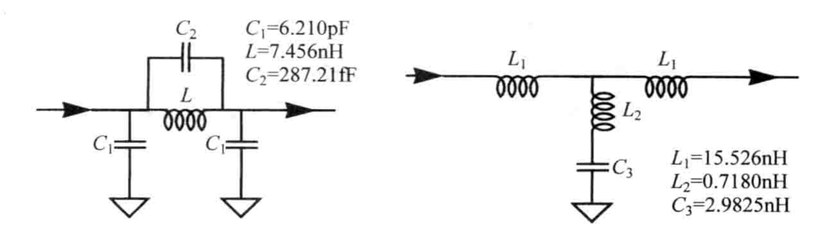


图 23.8　1GHz 集总低通椭圆滤波器原型（左边的是并联陷阱；右边的是串联陷阱）

在对下面的设计进行评估时，要记住集总原型给出了 1dB 的通带抖动，1GHz 的通带沿，以及开始于 3GHz 的阻带。在阻带中要求的最小衰减为 30dB。

就像从图 23.8 中可以看到的，右边电路的这个实现可以更好地转换为微带线支路形式。采用与之前相同的步骤，15.526nH 的电感是用 200Ω 的传输线实现的，其在截止频率处的电长度大致为 27.947°，而 0.718nH 的电感是用电长度大致为 1.292°的传输线实现的。类似地，2.9825pF 的电容用 15Ω 的传输线实现，其在截止频率的电长度大致为 13.465°。这个滤波器的仿真结果在下面两张图中给出。

这些仿真显示出通带误差直到 1.05GHz 都是低于 1dB，非常接近集总原型的 1GHz 值。在 3GHz 阻带边的衰减略低于 40dB，大大超过 30dB 的规定。同时，椭圆滤波器阻带陷波的特性一直到略高于阻带边是十分明显的[⑨]。参见图 23.9。我们的确预计频响在阻带中多少会回升些，但是这个分布式版本没能满足在频率高于 4.3GHz 上的 30dB 衰减的要求，这与集总原型不同[⑩]。

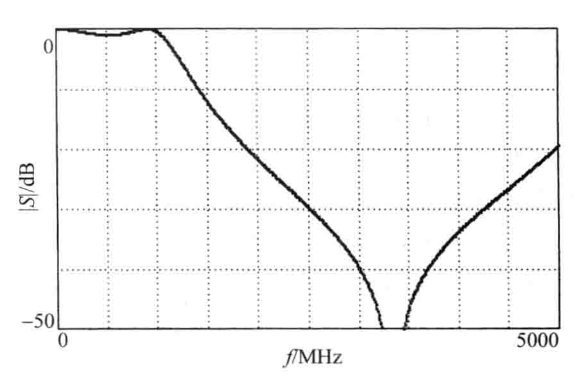


图 23.9　椭圆滤波器的支路版本的通带-阻带过渡详图

在一个更广的频率和衰减范围之内（如图 23.10），对频率响应的仔细观察进一步揭示了集总椭圆滤波器和对应的带支路传输线版本的区别。该滤波器的高频衰减很令人失望。同样地，对于任何设计，这样的在集总和分布实现之间的不同都是需要考虑到的。例如，如果需要更好的最终衰减，椭圆滤波器可以和另外一个阻带特性很好但在通带到阻带的过渡特性差的滤波器级联。通过这样设计的两个滤波器，其中每一个都被优化到不同组的目标，

⑨　再一次，有限大小的传输线的 Q 值不会导致有无穷深的陷落频点，而且也会引起插入损耗与其他的非理想行为。

⑩　但是，客观来说，（集总）元件的寄生参数能够引起集总滤波器同样非理想地工作。

可以得到一个具有所有希望得到特性的组合。

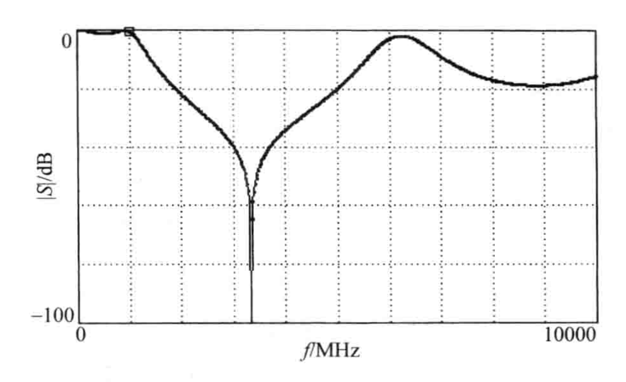


图 23.10　支路椭圆滤波器在大的频率和衰减范围内的幅频特性图

除了上面这些基本的设计考虑，对于这类滤波器的版图设计在高频下会比较令人头痛，亦即，在高的频率下具有高阻抗的电感性主干路的长度会缩减到这样小的程度，以至于宽的、电容性的支路会和相邻的结构相碰到。即便是在直接的相碰发生之前，线与线之间的耦合效应就可能以不可预计的方式改变传输函数。在这种情况下，需要通过使用比最小值更宽的线来加长主干路，这样做要付出的代价是主干路的阻抗特性多少要差些。类似的，可以用比最大宽度小的线来实现支路电容，这样做可能会是有帮助的。由于所有这些策略都有各自的折中选择，要有思想准备来遇到在设计这类滤波器需要的痛苦的迭代过程，同时不能保证设计目标一定能得到满足。

23.2.4　等比传输线滤波器

从公式(2)和(3)中，我们看到传输线长度和特征阻抗都可以用作自由度；你可以改变其中任一个或改变两者来产生希望得到的电感和电容。阶梯阻抗滤波器带有任意性地使用了传输线的两个固定的、极端的阻抗，而根据需要来改变其长度。一个互补的(但依旧具有任意性的)替代方法则是固定了传输线长度而根据需要来改变传输线的阻抗。由于所有的传输线段都是等长的，这样产生的滤波器称为等比传输线[11]。就如在阶梯阻抗滤波器的情形一样，末端短路的短传输线段被用来实现电感，而末端开路的短段被用来实现电容。在 Richard 最初对方法的描述中，短意味着为在截止频率处波长的 1/8：集总原型中每个电感和电容因此被 $\lambda/8$ 长的传输线替代，而用改变传输线的特征阻抗来得到所要的元件值。由于传输线的线段的长度是相同的，这样得到的滤波器响应在频率上具有理想的周期性(这与典型的阶梯阻抗或带有支路的滤波器的响应不同)，也可以被看作是集总原型混叠(aliasing)响应的结果。不用费太多努力，就能够推断出这个频响每隔 $4f_c$ 会重复，这里 f_c 为传输线长度为 $\lambda/8$ 时对应的频率。那些不可避免的凹角模式通常被认为是不希望的，但是它们有时也被利用(比如在 23.2.5 中描述的半波长滤波器)。有时候也会发生这样的情

[11] P. I. Richard, "Resistor-Transmission Line Circuits"(电阻器传输线电路), *Proc. IRE*, v. 36, February 1948, pp. 217-220。这篇文章将脚注 3 所引的 Cohn 描述的一些方法加以扩展，同时这篇文章是首次提供了一个对将集总滤波器变换成分布式滤波器的完整的理论分析框架。Richard 这个人的特别之处是他因为这个工作在两年内就拿到了哈佛的博士学位，尽管他没有首先拿到学士学位。

况，即在某个频率范围内，通过混叠会在通带和阻带之间出现比期望还要陡峭的过渡。Richard 转换的主要优点是对这些凹角模式的可预测性。不需要复杂的仿真就能够确定到它们在何处出现。除此之外，等比传输线滤波器并没有什么特别之处。事实上，固定了长度使得我们少了一个重要的自由度。考虑到实际宽度的限制，失去这个自由度(即调整线段的长度，译者注)可能会在实践中使得我们无法满足设计目标。在这种情况下，考虑采用带支路的滤波器作为一个替代办法是有价值的。

由于在该方法中 $\lambda/8$ 传输线的重要性，有必要研究它们的某些性质：

$$\frac{Z}{Z_0}=\frac{\dfrac{Z_L}{Z_0}+\mathrm{j}\tan\dfrac{\pi}{4}}{1+\mathrm{j}\dfrac{Z_L}{Z_0}\tan\dfrac{\pi}{4}}=\frac{\dfrac{Z_L}{Z_0}+\mathrm{j}}{1+\mathrm{j}\dfrac{Z_L}{Z_0}}=\frac{\mathrm{j}+\dfrac{Z_L}{Z_0}}{1+\mathrm{j}\dfrac{Z_L}{Z_0}} \tag{18}$$

Richard 选择 $\lambda/8$ 的智慧现在变得清楚了，这是因为那些正切项变为 1，这样就使得公式变得相当简单。当这样的传输线终断为短路，归一化的阻抗就简单地为 j，即纯电感性的。当终断开路，阻抗为 $-\mathrm{j}$(纯容性)。因此 Richard 方法的实现包括将电容用开路的 $\lambda/8$ 传输线替代，而电感对应于短路的传输线。遗憾的是，对微带线实现短路，并没有实际的方法可以做到。幸运的是，我们可以利用传输线的行为来将串联电感器转换为并联电容器。通常被称为广义科罗达(Kuroda)等效，这样的转换使我们避免了综合接地的元件的需要，进而使得滤波器容易用微带线来制造[12]。正式地说，一共有四种广义科罗达等价，但是由于互易性其中一半是重复的。剩下的两个科罗达等价，我们只用一个来完成所寻求的变换。那些支路的特征阻抗是设计自由度，它们可以被用来改变缩放比以得到更为实际的设计。参见图 23.11，那里在电感和电容器旁边标出的阻抗分别是用于实现这些元件的终端短路和开路 $\lambda/8$ 传输线的特征阻抗。

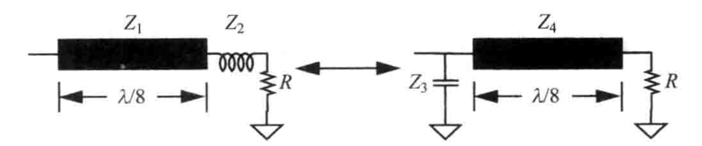


图 23.11　用于推导广义科罗达等效中最有用的几个等效电路

可以通过确定输入阻抗，然后令它们相等来推导相关等价性。对于用串联 LR 阻抗作为负载的传输线，

$$Z_{\mathrm{in1}}=Z_1\frac{\dfrac{R+\mathrm{j}Z_2}{Z_1}+\mathrm{j}}{1+\mathrm{j}\dfrac{R+\mathrm{j}Z_2}{Z_1}}=Z_1\frac{R+\mathrm{j}(Z_1+Z_2)}{(Z_1-Z_2)+\mathrm{j}R}=\frac{R+\mathrm{j}(Z_1+Z_2)}{\left(1-\dfrac{Z_2}{Z_1}\right)+\mathrm{j}\dfrac{R}{Z_1}} \tag{19}$$

类似的，对于另一个电路：

[12] 每一本标准的微波工程教科书(以及大量的文章)都会提到这些等效变换，但是几乎没有一处是给出其来源的。第一个提到这些变换的出版物很明显是 H. Ozaka 与 J. Ishii 的"Synthesis of a Class of Stripline Filters"(一类微带线滤波器的综合)，*IRE Trans. Circuit Theory*, v. 5, June 1958, pp. 104-109。Ozaki 与 Ishii 指出是 Kuroad 作为他的 1955 年的博士论文提出了这些等价性。很明显，Kuroda 本人从来没有在英文论文中发表过这些结果。

$$Z_{\text{in2}}=\left[Z_4\frac{\dfrac{R}{Z_4}+\mathrm{j}}{1+\mathrm{j}\dfrac{R}{Z_4}}\right]\parallel\frac{Z_3}{\mathrm{j}}=\frac{\left[Z_4\dfrac{R+\mathrm{j}Z_4}{Z_4+\mathrm{j}R}\right]Z_3}{\mathrm{j}\left[Z_4\dfrac{R+\mathrm{j}Z_4}{Z_4+\mathrm{j}R}\right]+Z_3}$$

$$=\frac{[Z_4(R+\mathrm{j}Z_4)]Z_3}{\mathrm{j}[Z_4(R+\mathrm{j}Z_4)]+Z_3(Z_4+\mathrm{j}R)} \tag{20}$$

这可以简化为

$$Z_{\text{in2}}=\frac{[R+\mathrm{j}Z_4]Z_3}{\mathrm{j}[R+\mathrm{j}Z_4]+Z_3\left(1+\mathrm{j}\dfrac{R}{Z_4}\right)}=\frac{R+\mathrm{j}Z_4}{\left(1-\dfrac{Z_4}{Z_3}\right)+\mathrm{j}\dfrac{R}{Z_3}\left[1+\dfrac{Z_3}{Z_4}\right]} \tag{21}$$

令两个输入阻抗表达式中的对应项相等，得到：

$$Z_4=Z_1+Z_2 \tag{22}$$

$$\frac{Z_4}{Z_3}=\frac{Z_2}{Z_1} \tag{23}$$

$$\frac{1+\dfrac{Z_3}{Z_4}}{Z_3}=\frac{1}{Z_1} \tag{24}$$

我们希望看起来束缚过度的方程组（我们只寻求 Z_3 与 Z_4，但有三个方程）不包括冲突。的确，更仔细地考察揭示了公式(24)中不包含前面两个公式中没有的信息，所以并没有问题。

求解公式(23)得到 Z_4，使其与公式(22)相等，而后求解 Z_3，得到：

$$Z_3=Z_1\left(1+\frac{Z_1}{Z_2}\right) \tag{25}$$

因此，公式(22)与公式(25)描述了我们将用到等价性中的一个。

也可以从前面的方程组中求解 Z_1 与 Z_2 的公式，得到同一个等价性，但是次序是倒过来的（如果以并联电容开始并且想将其转换为串联电感，这样做是有用的）：

$$Z_1=Z_3\parallel Z_4 \tag{26}$$

$$Z_2=\frac{Z_4^2}{Z_3+Z_4} \tag{27}$$

公式(22)和公式(25)～(27)放在一起，它们描述了图 23.12 中所示的等价性。同样，所指出的阻抗是那些用来实现这些元件的传输线段的特征阻抗。

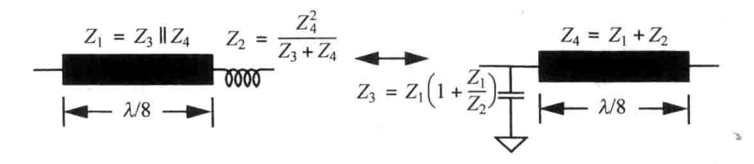


图 23.12　有关的科罗达等价性

我们现在利用这个能力来进行串联电感和并联电容的转换以实现我们已经将其转换为各种分布形式的恒 k 滤波器。我们一直在用的集总原型为七阶滤波器，其两端口是并联电容。然而，如果有偶数个电感，那么所需的转换次数要少很多（中间的那个电感需要额外的转换，因而带来复杂性；读者可以尝试奇数个电感的情况，来发现那时的情况）。因此，使用

端口为串联电感的原型形式作为替代(但是完全等价的)。电容为 6.4662pF 而主电感为 15.915nH(端口处电感为该值的一半)。

作为惯例,用归一化的元件值进行设计,而直到最后阶段再进行反归一化以简化中间的计算。这样做,又减小了产生误差的可能性,因此继续使用这种方法。

首先找到那个实现主串联电感传输线的阻抗:

$$Z_2 = \omega L = 100\Omega \tag{28}$$

或归一化值为 2。端口处的电感可以用其阻抗为这个值一半的传输线来实现。我们也给出实现电容的传输线的阻抗为

$$Z_C = 1/\omega C = 25\Omega \tag{29}$$

其归一化值为 0.5。

在用科罗达等价进行转换之前,滤波器看上去如图 23.13 所示。应用这些等价原理要求一个 $\lambda/8$ 的传输线段连接到每个需要转换的元件,但是常规的集总原型(包括图 23.13 所示的)并不能满足这个基本的要求。解决的方法是注意到我们可以自由地在滤波器输入和输出端口加上任意长度的传输线,只要这些添加的传输线阻抗为 Z_0。$\lambda/8$ 长的线段在文献中也被称为单位元件,添加这些额外的单位元件完全不会改变滤波器响应,但是这的确使得我们能够使用科罗达关系来将那些端口元件进行转换。此外,每次应用科罗达等价变换都会产生一个新的 $\lambda/8$ 传输线段,这又可以与另外一个需要变换的元件成对,如此等等。可以这样将变换从端口逐个向中心进行,直到在滤波器中心的那些元件都被转换成并联形式。然后,如果需要的话,可以反过来将变换向滤波器的输入和输出端口进行,以剩下的串联元件(重新)转化为并联的支路。

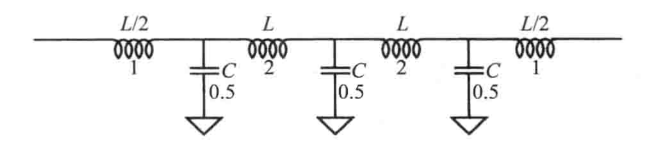


图 23.13　科罗达转换之前的滤波器(所有支路都是在截止频率时为 $\lambda/8$ 长)

这样,在我们这个特定的例子中,第一步是将端口的电感器通过在每一端添加一个单位元件(unit element,UE)转换为并联电容器,如图 23.14 所示。以同样的方法继续进行下去,就产生了一系列如图 23.15 所示的等效滤波器。

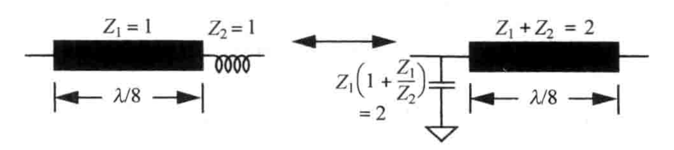


图 23.14　对举例的滤波器端口电感器应用科罗达等价变换

为了结束所有串联电感的消除,我们需要进行一个最后的变换以将两对外侧串联电感转换成为等效的并联电容(这些转换不一定要同时完成,但是这样做可加快变换速度)。完成了相关的转换后,我们最终得到了最后的设计参见图 23.16。

最后一步是将元件值反归一化以得到实际的传输线阻抗,然后再对设计进行仿真以确

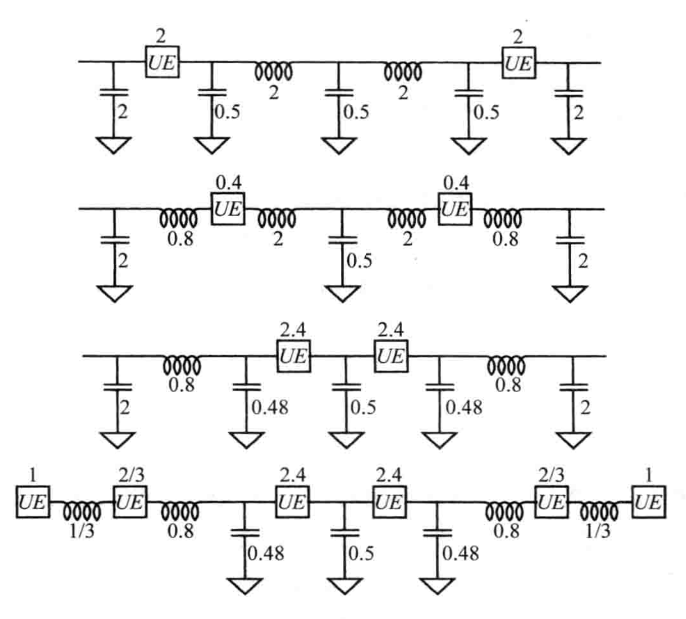


图 23.15　反复应用科罗达等价变换而产生的滤波器系列

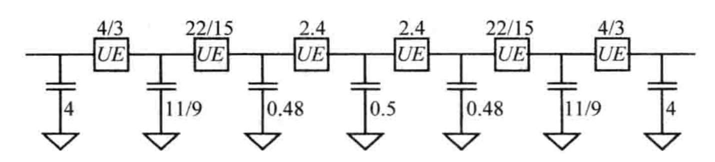


图 23.16　所有转换完成之后的最后得到的经归一化了的七阶低通滤波器

定在变换的过程中有没有引入错误(特别是考虑到有这么多的转化过程)。在计算了反归一化后的传输线阻抗之后,一个初步的滤波器版图如图 23.17 所示。支路的中心对中心的距离与支路本身的长度相同(尽管这个手绘的粗糙的图或许没有完全显示出这一点)。

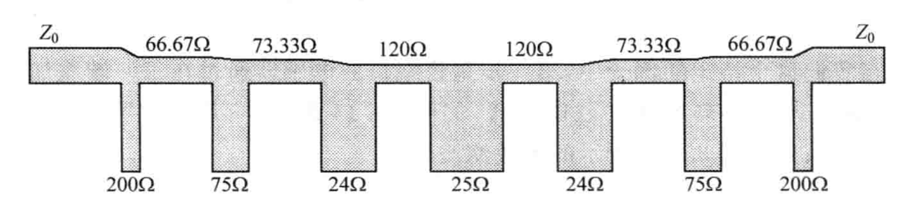


图 23.17　一种可能的 1GHz 带支路的微带线滤波器版图(没有按照比例绘制)

注意,这个演示性的版图仅仅通过改变上边界的位置来调整主干路阻抗。这样做虽然不一定在电性能上为最优,但是因为它的简单,这种选择几乎无一例外地被采用。

对这个设计进行仿真得到的结果如图 23.18 所示,从中可以看到滤波器性能如所期望的。1GHz 的衰减为－9.5dB(相比如恒 k 集总原型的－10dB),而－3dB 频点发生在大约 917MHz(相比较于集总原型的 934MHz)。在较大的频率和衰减范围内作图(图 23.19)清晰地显示了滤波器频响的预期的周期性。

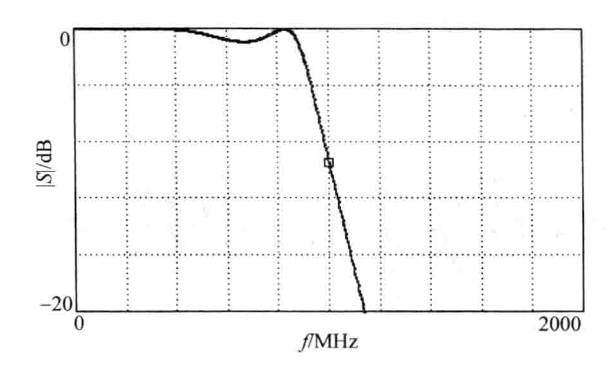


图 23.18　对经过反复应用科洛达等价变换得到的最终过滤器进行仿真的性能

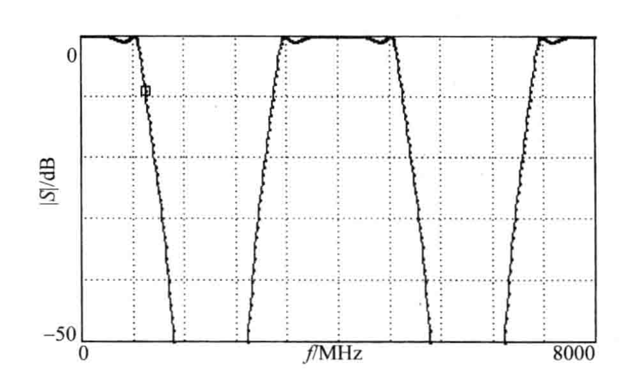


图 23.19　滤波器在更大的频率与衰减范围的频响性能

23.2.5　半波(凹角)“带通”滤波器

我们提到过具有等同于 λ/8 传输线的低通滤波器的频响每 $4f_c$ 重复一次，其中 f_c 为低通截止频率。值得强调的是这种周期性的凹形响应并不一定是无用或不希望的。作为说明这个问题的一个专门的例子，考虑一下利用图 23.19 中第一个失真的通带来产生一个带通滤波器。其带通的形状近似于它所基于的恒 k 低通滤波器的形状。很难说这个例子是唯一的：我们说过任何一个使用 Richard 的 λ/8 传输线的等同线低通滤波器会产生这种周期性频响，提供了很多应用带通滤波的机会。例如，在非常高的频率下，λ/8 传输线的尺寸会小到不能实现，那么或许就可以使用更高阶的通带。这种有意使用凹角响应模式的方法叫做过模(overmoding)。这对于毫米波和亚毫米波电路工作中特别常见，其中主模结构对于实际的制造工艺来说会太小。

在非关键的应用中，有一类带通滤波器用纸笔就可以轻松地设计。正如对阶梯阻抗低通滤波器设计起来不费劲，我们可以想象用基于元件长度等同的阶梯阻抗低通滤波器特例实现带通滤波器。注意如果线长在低通滤波器带的边沿为 λ/8，那么第一个带通响应的中

心频率对应长度为 $\lambda/2$。因此选择这些等同的元件长度为在通带的中心频率下 $\lambda/2$，见图 23.20。当使用这个结构为带通滤波器时，重要的一点是要谨记它同样也是一个低通滤波器这个事实，同时在任何设计中都要记住这件事。

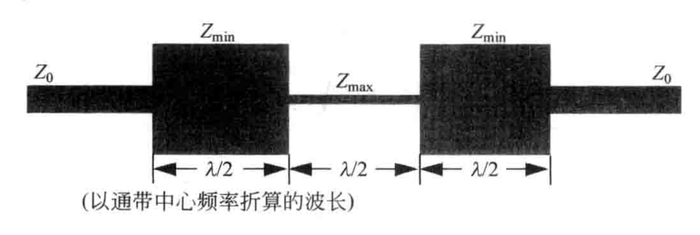


图 23.20　半波"带通"滤波器(等同阶梯阻抗版本)

对此类滤波器是如何工作的可以换一种角度来看，回顾一下 $n\lambda/2$ 长的传输线在输入端看进去的阻抗就是负载阻抗。对于我们这个特定的实现，这个条件对应着驱动点阻抗等于 Z_0，使得最大的功率传输到滤波器，并最终到达负载。对于滤波器单元电长度与半波长相距较远的频率，耦合到(或从其得到的)滤波器功率就没有那么有效了，这是因为在各种界面上存在着很大的阻抗失配。从这个描述中，可以看出这个滤波器的工作确实如带通滤波器那样，它具频率处在位置为周期性的通带，各通带中心所对应的频率为线长为半波长整数倍的那些频率。

对偏离中心频率的频率成分的抑制是通过使窄和宽滤波器段的失配尽可能大来实现最大化的。这个观察隐指每个宽半波长单元阻抗要越小，而每个窄单元阻抗要越高，就可以得到更好的阻带抑制，这个情形是与如何一个阶梯阻抗滤波器是一样的。阻抗比越为极端(即大)，阻带越深。此外，随着单元数目的增加，阻带的抑制能力也增加了。

图 23.21 所示的仿真结果是对应着图 23.20 的版图，采用了 15Ω 和 200Ω 传输线，在 250MHz 时长度为 $\lambda/8$，以使得第一个凹形的中心出现在 1GHz 处。频响中心在直流、1GHz 和 2GHz 在仿真中可以很明显地看到。为了在实际的滤波器中得到一个与这个频响接近的近似，就必须使高阻抗线适当长一些(增加的长度大约在介质厚度量级，更精细的估计可参考第 7 章)，以为 T 节缩短效应留出应有的余量。

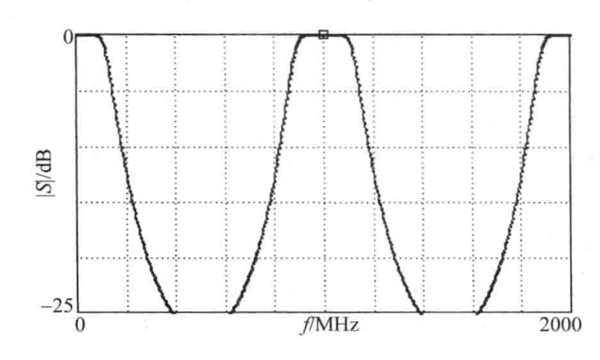


图 23.21　等比传输线阶梯阻抗滤波器的频响特性

作为一个最后的观察，我们应该注意到，尽管在该滤波器中使用了等同传输线，要求所有线段的长度相同并不是绝对的。作为一个极端的例子，将长度在 1GHz 为 $\lambda/4$ 的阻抗单

元与长度为 $\lambda/2$ 的高阻抗段连接到一起会在中心为 1GHz 的奇数倍处得到非常窄的通带，而在 1GHz 的偶数倍处得到宽的通带，见图 23.22。

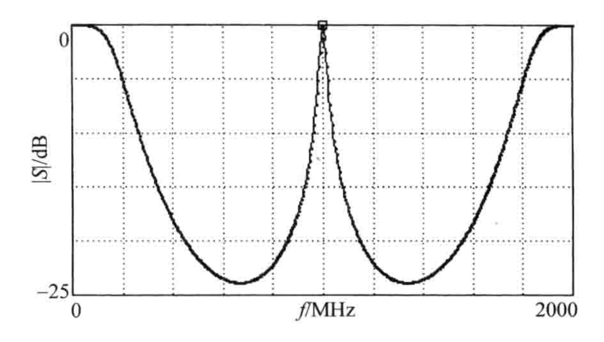


图 23.22　另外一个滤波器的仿真结果

诸如 Puff 仿真器对于快速评价改变长度、宽度和单元数目带来的影响是十分有用的。读者可以尝试使用 Puff 仿真器或其他合适的仿真工具来探索长度的改变如何影响频响形状，同时试着解释为什么会有这样的结果。

23.3　耦合谐振器带通滤波器

半波“带通”滤波器利用一种寄生效应来得到带通响应，这或多或少可以说是设计低通滤波器时偶然得到的副产品。我们现在考虑如何设计能直接而不是碰巧提供带通特性的滤波器。

23.3.1　集总带通滤波器

如我们所见的，经典的集总带通滤波器由平行的振荡器与串联共振臂耦合组成(图 23.23)。回顾一下这个结构是从带宽和其他特性如期望的带通滤波器对应的低通原型演变而来的。在这个特别的例子中，低通原型有两个并联电容和一个串联电感。所有这三个元件和它们对偶元件成对在期望的中心频率上产生共振。就是说每个并联电容和适当的电感并联，而每个串联电感也与电容相串联。不幸的是正是前面最后一步要求的串联共振单元妨碍了从集总原型到微带形式的直截了当的转换。例如，一些与并行回路中电感并联的寄生电容是允许的，它们可以成为共振器的一部分。然而，串联臂中的电感必须有很低的寄生电容，因为这个电容不容易被很好地吸收掉。

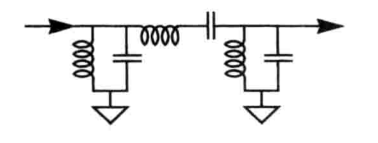


图 23.23　经典集总带通滤波器

对于集总和分布式带通滤波器，解决这个问题的第一步是意识到串联和并联振荡器彼此是相对仗的。就是说，如果我们计算其中一个网络阻抗的倒数，就会得到另外一个网络的阻抗(在一定的比例系数条件下以保证单位的正确)。具体地说，考虑一个串联 LC 振荡器的阻抗：

$$Z_S = sL_s + \frac{1}{sC_s} \tag{30}$$

而一个并联 LC 共振回路的阻抗是：

$$Z_P = sL_p \parallel \frac{1}{sC_p} = \frac{sL_p}{s^2 L_p C_p + 1} \tag{31}$$

如果我们计算串联网络阻抗的倒数，得到：

$$Z_{eq} = \frac{K^2}{Z_S} = \frac{K^2}{sL_s + 1/sC_s} = \frac{K^2 (sC_s)}{s^2 L_s C_s + 1} \tag{32}$$

其中已经引入了系数 K（具有阻抗的单位）来解决单位的问题。当然也可以将这个关系写作导纳形式，如下：

$$Y_{eq} = \frac{J^2}{Y_S} \tag{33}$$

其中再次引入一个合适的比例常数，J。为了使单位正确，J 需要有导纳的单位。

因为公式(31)与公式(32)形式上相同，可以看到我们的确能将串联谐振器转化为并联谐振器（反之亦然），前提是我们有能将阻抗取倒数的网络来做这件事。令对应项相等，可以推导出所需要的设计关系：

$$K = \sqrt{L_p / C_s} = \sqrt{L_s / C_p} \tag{34}$$

$$L_s C_s = L_p C_p \tag{35}$$

这样，将两个谐振频率设成相等，然后选择合适的转换因数 K（或 J）来满足公式(34)。如我们将要看到的，存在着一个额外的自由度，可以用来帮助在实际可行的范围内选择元件参数值。当然，所有这些是以我们假设有阻抗逆变器（impedance inverters）为前提的。在此先搁置如何实现阻抗或导纳求逆器（admittance reciprocators）（两者一起统称为阻纳逆变器，immittance inverters[13]）的问题，考虑它如何解决了串联谐振器的实现问题。如果将每个串联谐振器置换为两个逆变器中间夹的并联谐振器，从该组合的两侧看上去就成了串联谐振器。因此可以仅用并联谐振回路就可以实现带通滤波器。

例如，可以将图 23.23 中的电路转换为一个全并联谐振器等效电路，如图 23.24 所示。因此，不需采用两个并联谐振回路及一个串联谐振器，现在用三个并联谐振回路加两个阻抗求逆器来实现。

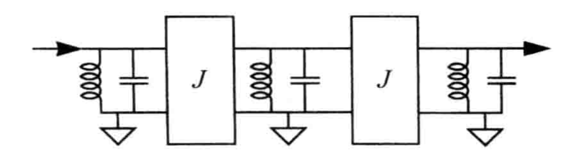


图 23.24　使用阻纳逆变器的全并联谐振回路带通滤波器

为了完成整个设计，需要考虑实现这些求逆器。实际上已经见过一种阻纳求逆器的形式：$\lambda/4$ 传输线。当然，仅仅在特定的频率下 $\lambda/4$ 条件得以满足使得这种求逆器并不是一个宽带元件。更受限制的是在相对小范围的实际可实现的特征阻抗限制了这种元件作为求逆器。尽管这些局限性，它偶尔依然是可以采用的，尽管这一选择经常被忽视了。

可以通过使用许多种集总网络的任一个来克服这个 $\lambda/4$ 线的某些局限性。阻抗和导纳

[13] Immittance 这个词是一个混成词，来源于将 impedance 与 admittance 这两个词合并在一起。这个词经常被误拼为 immitance。

求逆器的例子都在图 23.25 给出。如例子中所见，J 和 K 分别为导纳和阻抗的缩放系数。当然，转化阻抗的网络也必然能转换导纳，所以事实上两者之间并没有本质的区别。区别多少是人为的。因此，对于阻抗求逆器，

$$Z_{\text{eq}} = K^2/Z \tag{36}$$

而对于导纳逆变器有

$$Y_{\text{eq}} = J^2/Y \tag{37}$$

二者之间依如何方便来选择。

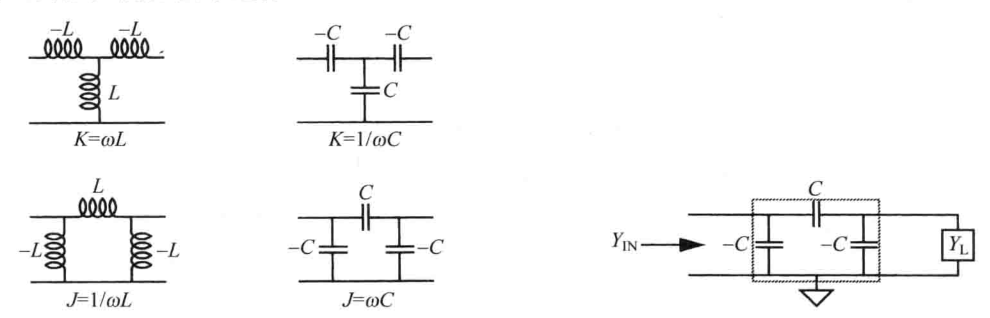


图 23.25　常用的阻抗（K）和导纳（J）逆变器

图 23.26　电容性导纳逆变器

我们将会看到电容的 π 网络是一个带通滤波器的特别有趣的例子，因此来验证一下它的导纳逆变性质。从图 23.26 中很清楚地看到输入导纳为：

$$Y_{\text{IN}} = \left[\frac{1}{Y_{\text{L}} - sC} + \frac{1}{sC}\right]^{-1} - sC = \frac{(Y_{\text{L}} - sC)sC}{sC + (Y_{\text{L}} - sC)} - sC = -\frac{(sC)^2}{Y_{\text{L}}} \tag{38}$$

这个网络因此的确是一个阻纳逆变器，其中特征导纳 J 等于电容性导纳。

我们将验证图 23.25 中 J 和 K 的其他表达式的任务留给读者作为练习。从这些表达式可以清楚地看出对于这些特定的网络，逆变是随频率而变化的。尽管如此，对于很多窄带（例如 10%～20%带宽）带通滤波器，这种求逆变换在通带内是足够平坦以可以用来实现际的滤波器。另外，由于电感和电容逆变器在频率上关系正好相反，交替使用它们能够以 2 或更大的倍数增加整体的转换带宽。

对这些网络的仔细观察无疑会使读者想到负电感和电容。由于逆变器从来不会单独使用，因此没有必要去设计一个负元件。通过对特定的结构选择合适的逆变器，那些负元件总是可以被余下的网络所吸收，在整个网络中它们仅仅减少了已存在的正电容或电感的数值。这个主意可以用一个特别的例子来最好地说明，因此选择一个如图 23.24 中所示的电路使用的电容性阻抗逆变器来得到图 23.27 中的电路。

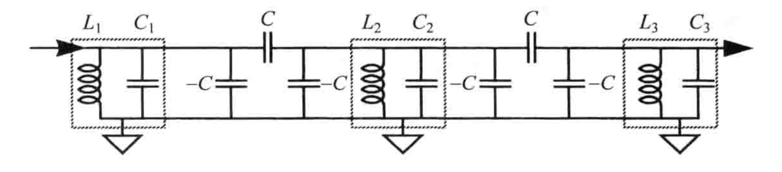


图 23.27　在合并电容之前的使用 J 求逆器的全并联谐振回路带通滤波器

作为第一步，注意到在用线框强调的并行振荡器都是调谐到同一频率，这是作为低通到带通转换的直接结果。值得把这个事实记在心上，因为将被证明不仅在我们考虑调谐实际

的滤波器的实验方法时是有价值的，而且在我们尝试将这个电路转变为分布式实现也是如此。

接下去的一步是将槽路电容和 J 逆变器的(负)电容相合并。这个过程如图 23.28 所示。经过并入负电容之后，所有的网络电容都是正的，如前所述的那样。尽管谐振器的谐振频率不再相同，只要将包围给定的电容周围的谐振器短路就恢复了“丢失”的电容，从而就恢复了谐振频率。这种不变性指出滤波器的调整对齐过程包含逐个将谐振器短路，以允许单个谐振器的调谐。同样地，当考虑调谐滤波器的实际方法时，会展开讨论这个想法。

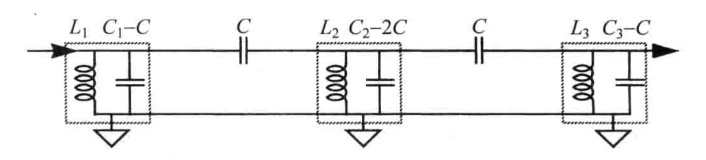


图 23.28　在将电容合并之后，采用 J 逆变器的全并联谐振回路带通滤波器

一个详细的例子

为了详细说明设计过程，让我们考虑如何来实现一个电容性耦合滤波器。从一个低通原型开始，应用低通到带通的变换，然后采用阻纳逆变器实现全并联谐振器。

假设目标是设计一个从 950MHz 到 1.05GHz 的带通滤波器。随意地从选择一个三阶 1dB 抖动 Chebyshev 低通原型开始，其抖动带宽为 100MHz[⑭]，参见图 23.29。这个滤波器中元件的数值如表 23.1 所示。

图 23.29　低通三阶 1dB 抖动 Chebyshev 原型滤波器 (100MHz，$Z=50\Omega$)

表 23.1　100MHz，1dB Chebyshev 低通滤波器原型的元件数值($Z=50\Omega$)

L	C
79.11nH	64.42pF

接下来，通过使每一个元件谐振于上下截止频率的几何平均值震荡来得到一个带通版本。通常不用担心算术和几何平均值的微小差别，因为对于这个 10% 带宽的例子，两者的差别小到可以被忽略不计。然而，为了使整个过程说明得准确，在这里做这样的区分，同时继续保留比实际需要更多的有效数字位。因此在这里，得到了谐振发生在 998.75MHz，参见图 23.30。这个带通滤波器的元件数值在表 23.2 中给出。

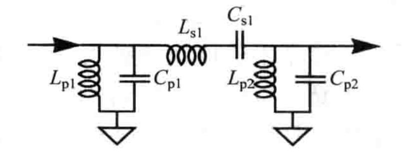


图 23.30　原型的低通滤波器 (100MHz 带宽，998.75MHz，$Z=50\Omega$)

⑭　我们有意选择了一个多少有点大的抖动以使得一个定值的抖动可以在幅度响应的图上清楚看出。实际情况下，遇到的抖动值更可能是十分之几个 dB，或更小，除非是在一些不重要的应用中。

表 23.2　100MHz 带宽，1dB 抖动，1GHz 带通滤波器元件数值（$Z=50\Omega$）

L_{p1}, L_{p2}	C_{p1}, C_{p2}	L_{s1}	C_{s1}
394.242pH	64.42pF	79.11nH	320.993fF

只是作为参考，带通滤波器原型的响应在图 23.31 中给出。明显地看出 1dB 的通带抖动，通带 100MHz 的带宽，其中心频率是在 998.75MHz 很好的近似。仔细看一下幅度的响应可以得出 −1dB 带边上限和下限的确是在 950MHz 和 1.05GHz，一如我们所设计的。

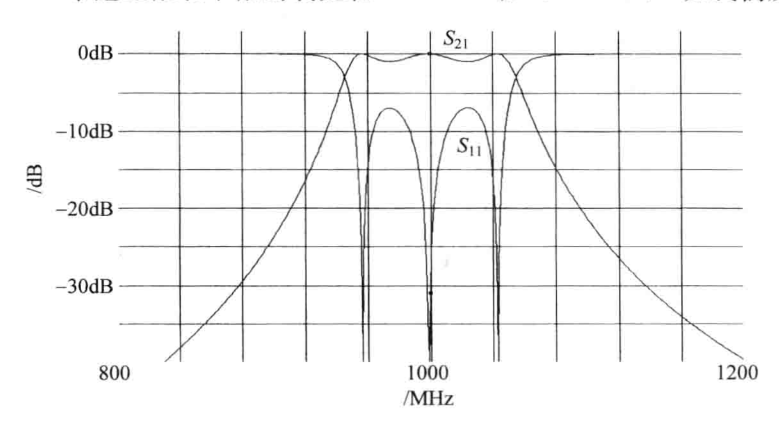


图 23.31　原型 Chebyshev 带通滤波器幅度和反射系数的频响（RFSim 99）

再下一步是采用阻纳逆变器来用并联谐振器替代串联谐振臂。在许多种逆变器的选择中，考虑到我们已花了很多的时间在理解电容性 π 网络上，同时也期望将集总设计转化为微带实现，我们选择了电容性 π 网络。

在将串联谐振支路变换为并联等效电路时，我们有一个限制条件不足的问题，因为当从原先的双元件串联谐振网络产生一个等效电路时，可以选择三个变量（并行 LC 网络元件参数，也包括 J 的值）的自由度。任意地选择转换后回路的电感（L_p）值等于其他的电感（在这个例子中是 394.242pH）[15]。这种选择使得必须要采购、设计与表征的不同的元件数目减少（这是对用分立元件实现而言）。使电容在中心频率谐振就产生了与其他两个并联谐振回路电容相同的回路电容，C_p（64.42pF）。

有了这些元件值，逆变器的特征电纳 J 因此设置在

$$J=\sqrt{\frac{C_{s1}}{L_p}}\approx\sqrt{\frac{320.993\text{fF}}{394.242\text{pH}}}\approx 28.534\text{mS} \tag{39}$$

这对应于 35.05Ω 的特征阻抗。所需的逆变器的电容 C 可以很容易地计算得到为

$$C=\frac{J}{\omega_0}=\frac{28.534\text{mS}}{(2\pi\cdot 0.998\,75\times 10^9)\text{rps}}\approx 4.547\text{pF} \tag{40}$$

参见图 23.32。在将元件合并之后，我们得到滤波器的最终形式，如图 23.33 所示。

作为一个验证步骤，考虑电容性耦合滤波器幅度和反射系数频响的仿真（图 23.34）。

---

⑮ 因为并不是所有的低通滤波器原型具有相等的电感，因此并不是所有从这些原型推出的带通滤波器有相同的电感。这里选择的特别例子正好是一个例外。

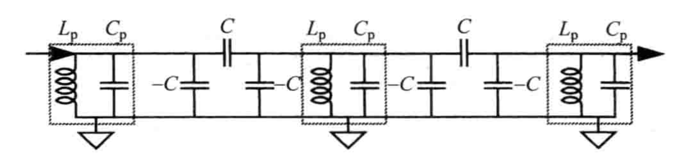


图 23.32　归并电容前的带通滤波器

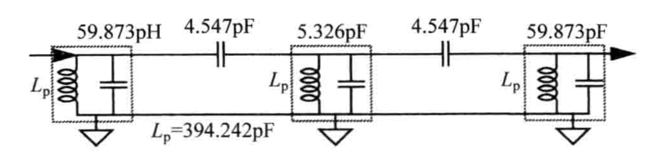


图 23.33　归并电容后的带通滤波器

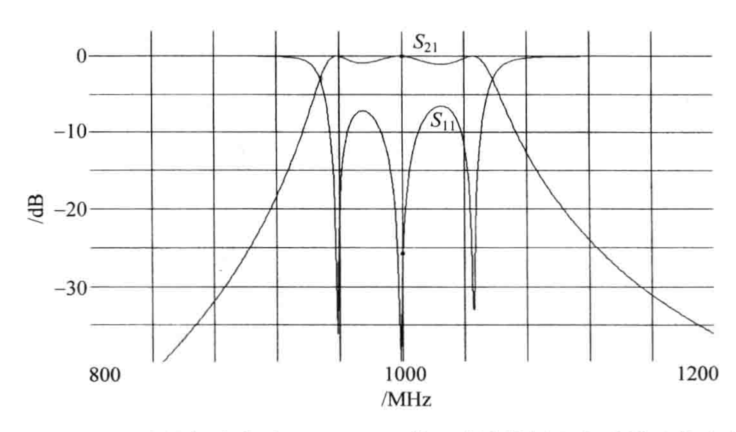


图 23.34　电容性耦合集总 Chebyshev 带通滤波器幅度和反射系数响应

很明显，频响与原型带通滤波器很相似，但是再仔细一看揭示了一些不同之处。由于逆变器的特征导纳随频率增加而增加，滤波器的频响并不那么对称：在高频时传输率更高些。这个特性在通带中几乎看不出来，尽管−1dB 带沿的上限和下限略微升高了一些(这里是到 953MHz 和 1.053GHz)且通带不再是等抖动高度了(抖动在高频处更大些)。对于分数带宽更大些的滤波器，这些误差可能会变得不能接受。在这种情况下，一个选择是使用稍微修正了的带宽(对于原型而言)和中心频率(对于综合后的带通滤波器)的定义，以减小由逆变器频率依赖特性造成的误差。在这种情形下，低通原型的带宽可以被设计成：

$$(\mathrm{BW})=\frac{f_0^2(f_u-f_l)}{f_l f_u} \tag{41}$$

这里 f_l 和 f_u 为最终的带通滤波器带的频率上、下限，其中心频率 f_0 现在由下式给出[16]：

$$f_0=f_l+f_u-\sqrt{(f_u-f_l)^2+f_l f_u} \tag{42}$$

对于小的分数带宽，这些公式收敛到我们一直在使用的较简单的公式(即，几何平均中心频率，与低通原型和带通滤波器带宽的相等)。对于这个特别的例子，这些公式告诉我们

[16] Seymour B. Cohn, "Direct-Coupled-Resonator Filters"(直接耦合谐振器滤波器), *Proc. IRE*, *February* 1957, pp. 187-196.

低通原型的带宽应该被设置在 104.3MHz，而带通滤波器中心频率大致需要设计到 996.26MHz。

同时，可测量的通带内抖动也稍微增加了。距离中心频率越远，这些差别更大。与原先的带通滤波器相比，电容性耦合滤波器在低的阻带中衰减更大，而在高的阻带中的衰减就小些。这些差别并不一定意味着严重的问题，但是它们的存在必须加以考虑与采取相应的措施。假若尝试去设计带宽要宽得多的滤波器，使用有频率依赖关系的阻纳逆变器引起的差别会变得明显得多。同样地，这些差必须被预测到并给予相应的考虑。比如说，如果原型低通滤波器刚刚能满足这些要求，电容耦合的滤波器版本就不能满足其中一个或更多的滤波器指标。

使用阻纳逆变器的带通滤波器设计总结

由于我们刚给出的设计步骤的推导分散在几页纸上，为了今后引用起来方便，将这些结果用菜单的形式简洁地总结出来是有帮助的。除了便于集总、电容耦合的带通滤波器的设计，我们将会看到这些公式对于一类重要的全分布式滤波器的设计也是直接有用的。

设计的第一步是产生一个集总低通原型，其通带和阻带特性（例如，抖动和带宽）就是那些带通滤波器所要具备的。出于符号一致性的考虑，假设原型滤波器归一化到 1rps 带宽与 1Ω 阻抗，电感和电容统一用变量 g_k 表示（注意：这里的 g 不表示电导）。用下标 p 和 s 分别表示并联和串联支路元件。

对于一个具有归一化带宽(BW)，阻抗为 Z_0，及中心频率为 ω_0 的带通滤波器，并联电容的值为

$$C_p = \frac{g_{kp}}{\omega_0 Z_0 (BW)} \tag{43}$$

与如下值的电感共振：

$$L_p = \frac{Z_0 (BW)}{g_{kp}\omega_0} \tag{44}$$

类似的，串联电感由下式给出：

$$L_s = \frac{g_{ks} Z_0}{\omega_0 (BW)} \tag{45}$$

与值为如下的电容共振：

$$C_s = \frac{(BW)}{g_{ks}\omega_0 Z_0} \tag{46}$$

每个串联谐振器可以被一个并联谐振器与两个阻纳逆变器的组合来替代，以实现全并联谐振回路的电路。随意地选择用导纳来表征阻纳逆变器，有：

$$J = \sqrt{C_s/L_p} = \sqrt{C_p/L_s} \tag{47}$$

这些谐振器的特征阻抗(Z_{0res})不需要和外部系统的阻抗相同。但是使它们相等的确简化了设计，因为可以消除在输入输出端口添加阻抗匹配单元的需要，这或许就是为什么大多数教科书没有明确指出额外的自由度的原因。然而，保留这个自由度可能是会有好处的，因此对于集总并联谐振回路我们会写作：

$$Z_{0res} = \sqrt{L_p/C_p} \tag{48}$$

逆变器导纳和并联谐振回路阻抗的乘积因此变为

$$JZ_{0\text{res}} = \sqrt{C_s/L_p}\ \sqrt{L_p/C_p} = \sqrt{C_s/C_p}$$
$$= \sqrt{\frac{(\text{BW})}{g_{ks}\omega_0 Z_{0\text{res}}} \Big/ \frac{g_{kp}}{\omega_0 Z_{0\text{res}}(\text{BW})}} = \frac{(\text{BW})}{\sqrt{g_{kp}g_{ks}}} \tag{49}$$

给定其他参数这个方程可以很容易被用来求解逆变器的导纳。注意公式(49)的最右端的那项是完全用归一化量来表示的。因此，乘积 $JZ_{0\text{res}}$ 可以容易地由给出的归一化带宽和低通原型元件数值计算得到。根据需要重复该计算，来将每个串联支路变换成并联支路，在每一步这样的变换中采用恰当的 $g_{kp}g_{ks}$。

当并联谐振回路阻抗不等于系统阻抗时，可以使用阻纳逆变器来提供所需的阻抗变换。与 $\lambda/4$ 传输线阻抗匹配单元类似，可以将逆变器的特征阻抗设成与需要匹配阻抗的几何平均值一样来得到希望的结果。因此，如果原型低通滤波器从归一化值 g_1 的串联电感开始，则匹配至 Z_0 的阻抗需要：

$$J = \frac{1}{\sqrt{Z_0\omega_0 L_1}} = \frac{1}{\sqrt{Z_0\dfrac{g_1 Z_{0\text{res}}}{(\text{BW})}}} = \frac{1}{\sqrt{Z_0\dfrac{mg_1}{(\text{BW})}}} = \frac{1}{Z_0}\sqrt{\frac{(\text{BW})}{mg_1}} \tag{50}$$

其中 $m=Z_{0\text{res}}/Z_0$。当用公式(49)相同的形式表示，这个表达式变为

$$JZ_0 = \sqrt{\frac{(\text{BW})}{mg_1}} \tag{51}$$

如果需要其他合适方法的话，可以用 $\lambda/4$ 传输线来实现这些逆变器，而不是用集总网络。

注意阻抗变换器的存在也允许源和负载的阻抗可以不相等，也可以不同于 Z_0。因此，如果 g_N 为最终原型滤波器元件的归一化数值，那么有：

$$JZ_0 = \sqrt{\frac{(\text{BW})}{mg_N}} \tag{52}$$

尽管通常源与负载阻抗是相等的，但是当它们不相等时，这个自由度偶尔也是很有价值。

下一步是选择逆变器的拓扑结构并确定用于给出要求的 J 的每个反相器元件的数值。然后将逆变器的元件并入剩下的网络以消除负的元件数值。如前所述，交替使用电容和电感性的逆变器扩展了变换的带宽。当设计分数带宽(如，30%～40%)大的带通滤波器时，这个选择很有价值。

集总和分布式元件混合的带通滤波器

对集总元件逆变器的强调反而会容易忽略分布式逆变器，毕竟后者是我们开始考虑阻抗逆变这个问题上的初衷。因此对由三个集总谐振器与一对 35.05Ω 的 $\lambda/4$ 传输线构成的逆变器相耦合而组成的滤波器(参考图 23.35)频响进行讨论是很有指导意义的。由于这个传输线的特征阻抗是在实际可实现的范围之内的，也由于谐振器和我们这个特定的例子相同(而不是几乎相同)，这样的滤波器值得考虑，特别是因为它们有不错的性能。

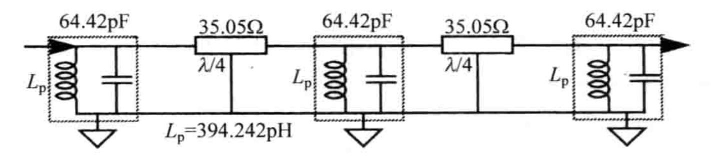


图 23.35　使用集总谐振器与传输线逆变器的带通滤波器

这个滤波器的幅度频响如图 23.36 所示。很明显，通带响应的确对应于具有 1dB 抖动的 Chebyshev 滤波器的响应。另外，在远离中心频率的地方，其特性比在用集总元件实现的逆变器的滤波器多少要好些。然而，1dB 抖动通带是从 954MHz 扩展到 1045MHz，因此在带宽上有所减小。同时，通带的抖动也减小了些，大致为 0.9dB。使用 Cohn 更为复杂的公式来计算中心频率与带宽，能够在很大程度上纠正这些不算大的问题。

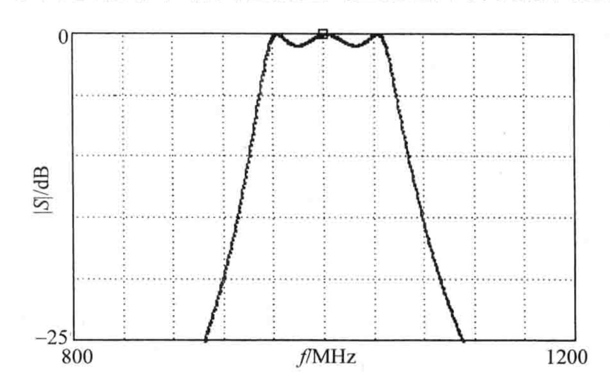


图 23.36　使用传输线耦合集总谐振器的带通滤波器频响

可以通过更换集总谐振器为等效传输线来得到完全分布式的带通滤波器。这样做在原理上是很直接的，因为可以将并联 L_pC_p 谐振回路替代为 $\lambda/4$ 传输线，这个传输线的特征阻抗只是与并联谐振回路的稍有不同[17]：

$$Z_{0\text{stub}} = \frac{\pi}{4}\sqrt{L_p/C_p} \tag{53}$$

仅作为参考，当将一个串联 LC 谐振器用终端开路的 $\lambda/4$ 线替代时，对应的关系为

$$Z_{0\text{stub}} = \frac{4}{\pi}\sqrt{L_s/C_s} \tag{54}$$

这些关系式是由令反射系数和导纳在谐振频率附近处的斜率(即 dZ/df 或 dY/df)相等而推出的，而不是用在远离共振的频率处的斜率(参考 23.6 节)。

在我们的例子中，集总并联谐振回路的特征阻抗已经是一个较小的值了：2.474Ω，因此等效的末端短路的逆变器会具有更低的特征阻抗：1.943Ω。不用说，实现这样低阻抗的传输线实际上是不大可能的，但是这个问题可以通过用一个公共的系数来放大所有阻抗，而后在输入输出端口进行阻抗转换的缩放。唯一的限制是我们必须注意不能要求过于大的逆变器的特征阻抗。这里，我们非常幸运，将所有阻抗放大 5～6 倍给出谐振器和逆变器的阻抗恰好落在实际 FR4 可以实现的范围之内。这个例子比较特殊，因为经常会发生要求传输线阻抗超过可以达到的范围的好几倍。这个情况让我们更有动力去考虑替代的分布式谐振滤波器。其中的一个结构保留了相对低阻抗的分布式谐振器，但是将它们与集总电容耦合；而电容的可用值范围是很广的，使得在很多情况下的实现变得可能。事实上，或许是最常见的(从生产数量的角度上看)微波带通滤波器使用的正是这个架构。滤波器采用高介质常数材料能很好地缩小所需体积，因而更加有吸引力。很多在 20 世纪 80 年代中期之后生产的

---

[17] 对于非常粗略的手算，通常可以任意地忽略 $\pi/4$ 这个因子(这样公式记起来当然要容易些)，以及忽略对串联等效的 $4/\pi$ 因子。

手机包含了至少一种这类的滤波器。尽管这些滤波器不是平面式的，它们的盛行证明了这里给出简短叙述的必要性。

如图 23.37 所示，这一类中的典型滤波器(有时被称为整体滤波器)具有特别简单的结构。一个导电矩形空腔用介质材料填充，只有上表面不被金属覆盖。在这些滤波器中材料典型的介质常数为 10～100。这种结构使得实现紧凑型滤波器成为可能(比如，对工作频率为 GHz 低端，最大边长量级为 1cm)，大多数这些介质是压电的，那么需要注意麦克风效应，甚至会产生破坏性的高电压的可能[18]。好消息是介质常数通常温度系数小于 10ppm/K，这要归功于材料科学家几十年来的研究贡献。另外，这些电介质的损耗非常小，因此谐振器通常的 Q 频率乘积为几个 THz。

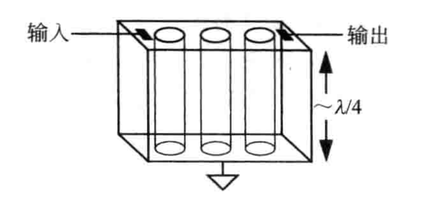


图 23.37 典型的陶瓷介质带通滤波器

每个末端短路的 $\lambda/4$ 谐振器是通过在介质材料中形成一个圆柱孔，并将圆柱孔内墙和底面镀上金属层来构造的。介电常数和谐振器的物理尺寸共同决定了谐振器的特征阻抗。相邻谐振器之间的耦合由它们相互间的距离控制，而这些谐振器到输入、输出端口的电容耦合则是通过将金属电极放置于终端谐振器一定的距离来提供。这样就得到了如图 23.38 所建模的结构。在手机中使用的典型单元的高性能可从它们的插入损失大致为 1～2dB 而观察到。

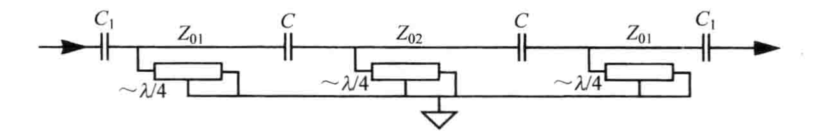


图 23.38 陶瓷介质带通滤波器近似模型

这些谐振器的特征阻抗略有不同，它们在长度上略小于 $\lambda/4$，这是因为用于将谐振器耦合在一起的导纳逆变器中的负的并联电容的作用的缘故。这些效应可以在此滤波器的全集总实现中容易看到。事实上，集总原型可以直接转换为图 23.38 的形式来得到精确的谐振器长度和阻抗。

我们看到对于基本的耦合谐振器实现可以有很多可能的不同版本。现在让我们回到这些版本中一个相关的子集，其在理论上是可能用平面工艺实现的。为了比较的方便，让我们仿真一个不实际的、全分布式滤波器(图 23.39)，而不做任何阻抗缩放。这样做了后，我们发现通带抖动和 1dB 抖动带宽比全集总电容耦合的实现依旧较小。将集总并联谐振回路用末端短路的传输线来替代在通带上引起了一个可以忽略的偏移，现在大致从 953MHz 扩展到约 1045MHz(如图 23.40)。为了纠正 8%的带宽误差，需要重复整个滤波器的综合过程，从带宽比最终设计目标大 8%～10%的低通原型开始。类似地，对中心频率的任何修正，也可以通过对用于从低通原型计算出初始的带通原型的中心频率进行适当的修改来实

[18] 电介质的组成可以是不同的，而且不断地在发生变化，但是它们通常包括铅、锆、钡、锶与钛。

现所需要的中心频率修正[19]。

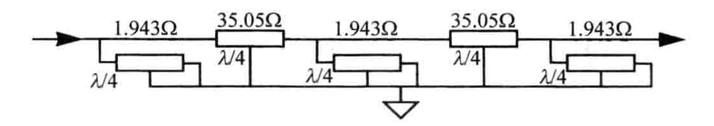


图 23.39　不实际的全分布式带通滤波器

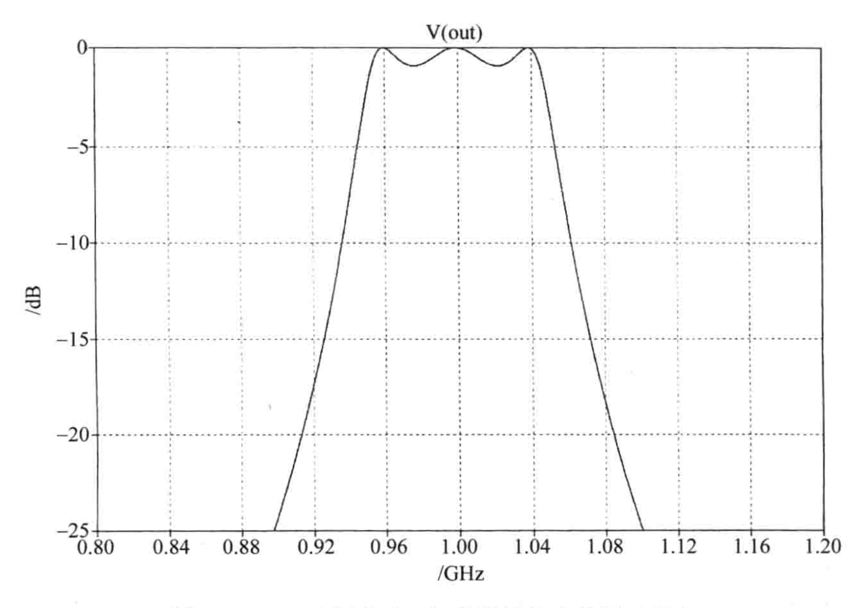


图 23.40　不实际的全 λ/4 带通滤波器的幅度频响

我们看到纯分布式的带通滤波器设计在原则上是可行的，但是经常会遇到十分不现际的阻抗量级，如果直接遵照上面给出的综合过程来做。为了发展一个能使我们规定最后结果处的谐振器阻抗（以此来保证阻抗值局限于实际的数值范围之内）的自由度的替代架构，需要拓宽我们对耦合产生的原因和效应的理解。将会看到，这种理解的拓宽带来的一个附带好处是校准（即调谐）这些滤波器的实验技术，以使得实际实现的滤波器会更好地近似那些激发它们的理论原型。

23.3.2　能量耦合和模式分裂

从高层次的角度来看，集总滤波器设计的目标是将传输函数的极点和零点放置在恰当的位置以得到期望的频响形状。这个重要的思想是所有现代集总滤波器的基础，无论是带通或其他类型。关于带通滤波器如何产生各种各样的频响形状的一个有用的描述是它们利用了所谓的模分裂，这是当两个或更多的谐振系统互相作用时发生的[20]。就是说，当两个相

⑲　Cohn 方程组能够减小，但通常不会消除相关的那些误差（典型地来说，在最后得到的滤波器的带宽的残余误差要比在中心频率上的误差大得多）。认识到应用这些方程依然产生误差，同时也看到现代模拟工具的可用性，采用我们已经使用的步骤是合理的，然后要在以后的过程中实施必要的调整。

⑳　实际上，这个机制也可以被引述来解释为什么，比如说，一根集总低通 RC 线的极点是类似地不相重合的。我们将集中在带通的情况，因为它对我们目前讨论的问题直接有关。

同的谐振回路用某种方式连接到一起，所得到的耦合系统的极点不会保持不变。正如我们马上要看到的，这种交互作用越强，引起的极点分开距离就越大。这个内在的原因与发生在耦合传输线上的偶数和奇数模阻抗分裂非常相似，这个比喻值得我们记住。

首先考虑两个相同、简单的 LC 谐振器，它们的电感器通过磁相互耦合在一起。之所以选择不那么熟悉的(至少对于现在学生来说)磁耦合，一方面是出于教学的原因(正是因为学生对此不是那么熟悉)，另外一方面是帮助强调将要推得的结论是非常具有一般性的。

将耦合电感器表征为一个变压器，进而将变压器建模为 T 形连接的三个电感，参见图 23.41。电感 L 为每个孤立谐振器中的电感。互感 M 为 L 的一个部分，取决于耦合的大小(及符号)决定。随着两个电感器的磁通量联接从 0～100%，耦合系数 k 大小的范围从 0～1[21]。在该模型中，当另外一个子回路开路时，左右两个子回路中的任一个的总串联电感为 $(L-M)+M=L$。

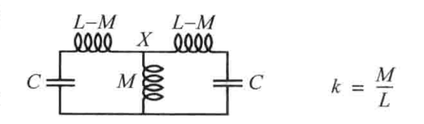


图 23.41　磁耦合的 LC 谐振器

为了找到所得到的四阶系统[22]的谐振频率，总是可以使用蛮力的方法：得到传输函数(这首先要定义输入输出端口)，然后求解分母多项式的根。这个方法是十分通用的，但是也十分繁琐的，特别是对于阶数高于 2 阶或 3 阶的网络。更糟糕的是，得到的结论与消耗的努力相比并不划算，因为得不到内在的认识。这里，这个网络碰巧是对称的，这样的一个情况几乎总是可以用来简化分析，并提高得到有用的内在认识的可能性。

首先回顾一下极点是什么。教科书称它们为传输函数分母的根，但它们更深层次的意义是它们是网络的固有频率。这个词的意思是：如果给系统一个初始能量，那么在没有进一步能量输入的情况下，系统状态在特征频率处发生演变，这些特征频率的值就是那些极点的值。系统状态可以以振荡的形式演变，意味着至少一些极点有非零虚部，也可以是对应于简单的普通指数和的行为，表明所有的极点为纯实数。通过谨慎选择初始状态可以仅仅激发所有可能模式中的一小部分，因而将困难的高阶问题转化为一些较为简单的低阶问题。非常巧妙(或者说有运气)的选择甚至可以得到激发单一模的情况，这样就可能通过求解最少数量的根来确定极点频率。

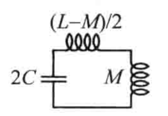


图 23.42　对于给定的共模初始条件的耦合 LC 谐振器的等效网络

我们可以用这个理解来设计一个简单的方法以找到耦合谐振系统的极点。首先，给两个电容器放入相等的初始电荷量，这样就提供了一个共模(偶数模)激发。不管网络之后如何演变，由对称性知道两个电容器上电压的变化应当是相同的。由于它们的电压总是相等，可以将这两个电容器短路而不会有其他影响，这样就得到图 23.42 中给出的网络。共模谐振频率因此对

[21] 我们必须强调负的 k 值并没有违反任何自然界的定律，取决于两个电感器的相对朝向，由在一个线圈里流动的电流在另一个线圈上感应出的电压可以是正的或负的。因此负的 M 值在物理上是可以实现的，这样就允许某些需要有负电感的网络的综合。事实上，Tektronix(美国的一家著名仪器制造公司)的几代示波器正是依赖于这样的电路元件(在电容性桥接的“T”形线圈)来提供垂直偏转放大器带宽的增加，就如在第 12 章中讨论过的那样。

[22] 尽管在这个网络中有 5 个储能元件，这个系统依然只有 4 阶，因为并不是所有的元件都是独立的。比如说，注意到只要描述三个电感中的两个中的初始电流就自动地决定了在第 3 个电感中的电流(根据基尔霍夫电流定律)，因此这三个电感实际上只贡献了两个自由度，从而对整个网络降了一阶。

应于一个简单的 LC 网络的谐振频率：

$$\omega_{\mathrm{cm}}=\frac{1}{\sqrt{[(1-k)L/2+kL]2C}}=\frac{1}{\sqrt{(1+k)LC}} \tag{55}$$

存在着具有这个频率的两个共轭虚极点，因此仅仅需要求解这个四阶网络的另外两个极点[23]。

因为共模的初始状态对于求解极点中的两个如此有成效，接下来尝试差分的初始状态自然显得很合理。特别是，如果一个电容的初始电压被设置在 V，而另外一个设置在 $-V$，那么由(反)对称性有，图 23.41 中节点 X 的电压一定保持为零，而不论系统状态以怎样的方式变化。因此，互感中没有电流流过，可以去掉(既可以通过短路也可以开路；两种方法必然会给出相同的结果)。去除那个电感就给出下面的差模共振频率(同样地，对应的两个极点是共轭纯虚数，其大小如下)：

$$\omega_{\mathrm{dm}}=\frac{1}{\sqrt{[2(1-k)L](C/2)}}=\frac{1}{\sqrt{(1-k)LC}} \tag{56}$$

既然已经找到了极点频率，让我们从中得到一些直观结论。首先考虑离零较近的值 k。在这种弱耦合的情况下，两个模频率几乎相同，因为两个回路几乎独立且相同。随着 k 增加，一个谐振频率增加而另外一个减小，发生模分裂。耦合越强，谐振频率的分裂越明显。这种特性应该并不陌生，在耦合传输线中我们就看到越强的耦合带来更明显的谐振频率分裂。正是插指的布局安排带来的强耦合使得兰格(Lange)耦合器有惊人的大带宽。

作为说明模分裂是耦合带来的必然结果，现在考虑电容耦合的对仗情况(图 23.43)。这里，我们选择将单个的谐振电容表达成为耦合电容的函数。也可以简单地将每个谐振电容记为有一个初始值 C，但这里的选择使得与电感耦合情况完全可以作类比，同时也将允许显式地确定一个电容 J 逆变器的存在。除了这些考虑，这也简化了解析表达式。在这个模型中，当另外一个回路短路(短路条件与研究电感性耦合时的开路为对偶关系)，每个回路的总电容为 $(C-C_c)+C_c=C$。正是由于这种对偶的特性，我们很快会看到，得到的公式是同构的。

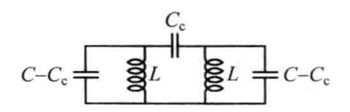


图 23.43　电容耦合谐振器

与分析磁耦合情况的方法类似，求解得两个模频率如下：

$$\omega_{\mathrm{cm}}=\frac{1}{\sqrt{(C-C_c)L}}=\frac{1}{\sqrt{(1-k)LC}} \tag{57}$$

$$\omega_{\mathrm{dm}}=\frac{1}{\sqrt{(C+C_c)L}}=\frac{1}{\sqrt{(1+k)LC}} \tag{58}$$

对于这些公式，我们看到耦合系数 k 的显式表达式为

$$k=\frac{C_c}{C} \tag{59}$$

与磁耦合情况相同，当用这种形式表达时，耦合系数不能超过 1(如果不允许出现负元件值)。更重要的是，磁和电容耦合给出了相同模分裂。这种机制很普遍，解释了一些看起来不相关的现象，例如半导体中能带的形成(这里，自由分立原子初始相同的模频率——能

[23] 我们是因为不存在着损耗这个事实来得出它们是纯虚数的结论的。因此，这个系统的能量始终保持为常数。

级——随着原子变得更近，形成晶体结构，而发生分裂）以及耦合弹簧-质量块系统振动模式。事实上，我们可以推而广之来说模分裂是由于能量的耦合。

从公式(57)和公式(58)中，应该可以清楚看到，通过测量两个模频率能在实验上来确定 k 的值。对于小的耦合值，模频率（归一化到它们的算术均值）的差相当准确地等于 k（事实上准确到二阶）[24]

$$\frac{\Delta\omega}{\omega_{\mathrm{am}}}=2\left[\frac{\sqrt{1+k}-\sqrt{1-k}}{\sqrt{1+k}+\sqrt{1-k}}\right]=2\left[\frac{1-\sqrt{1-k^2}}{k}\right]\approx k \tag{60}$$

对于滤波器（就像耦合线耦合器情况一样），0.5 的耦合系数被认为是很高的值。然而，即使对如此大的值，实际分数带宽大致为 0.536，与耦合系数的差别仅有 7%或接近的数值。对于更为典型的耦合系数，误差通常比器件容差造成的误差还要小——而且通常比实验上能够确定的精度还要小。出于这样的原因，很多应用说明和教科书称频率上归一化的分裂等于耦合系数，尽管并不是严格准确的。

如果读者要求比较苛刻，可以使用前面的公式推导出下面这个耦合系数的严格表达式，是用模频率的上下限 ω_1 和 ω_2 表示的：

$$k=\frac{\omega_2^2-\omega_1^2}{\omega_2^2+\omega_1^2} \tag{61}$$

可以应用这个关系来直接测量耦合。其中一种可行的实验过程是将一个信号发生器与一个谐振器松散地耦合（采用松散耦合是为了避免它自身的会诱发误差的模分裂），而后用测量仪观察被耦合的谐振器的响应（同样，耦合要松以避免干扰系统），在此同时扫描信号发生器的频率。作为一个很好的近似，耦合系数即为观察到的响应峰值差的归一化值。

为了将前面得到关于耦合的观察与我们已经研究过的电容耦合带通滤波器联系起来，首先重绘图 23.44 中的双谐振器系统。以这种方法绘制的原理图显式地揭示出将两个振荡器耦合在一起的三个电容就是前面部分所讨论过的用来产生电容耦合带通滤波器中的 J 逆变器。因此较强的耦合是与大的逆变器导纳联系到一起的。类似地，对磁耦合情况的研究表现出存在一个电感 K 逆变器，如图 23.25 的第一个网络。在这两个例子中，连接到逆变器每个端口的回路，再一次是在同一频率谐振。电容和电感耦合对模分裂是等同有效的这个事实再次强调了正是能量的耦合造成这种现象。尽管我们最初是在导纳逆变和集总耦合中观察到这个等同性，能量耦合的普遍性表明这个等同性不仅仅是限制在集总例子中的。遵循这个直观认识形成了一类完全分布式带通滤波器的基础，称为耦合线滤波器，这些滤波器中，磁场与电场都同时提供阻纳逆变和传输线谐振器之间的耦合。

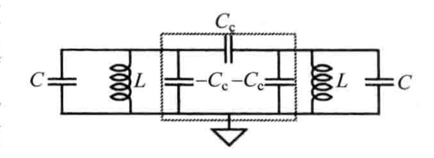


图 23.44　重绘的电容耦合谐振器

[24] 许多参考材料认为是归一化的带宽才精确地等于 k，但是可以请读者用几何平均值、算术平均值及单个的（未经耦合的）谐振频率（这是 Dishal 的最初的定义）中的任一个作为归一化因子（也存在着其他可能的选择）来验证这个说法。在所有这些情形下，归一化了的带宽多少与 k 不同。然而，幸运的是产生的误差是足够小而不会造成严重的实际影响。我们将要考虑的滤波器设计方法反正只是对（相对）带宽小于 20%才准确，对应的 k 值也是在这个范围。对于在这样的滤波器中的耦合的谐振器对，这里给出的近似相当精确地预测了模分裂。

23.3.3　微带线边缘耦合带通滤波器

分布式谐振器可以用很多方式耦合，不仅能给出带通响应，也能提供带阻和其他频响特性。对这类滤波器的研究紧跟二战之后开始得到重视，大多数形成现代微波滤波器设计基础的理论概念到了1960年都已成熟了。在那个多产年代的很多进展在 Matthaei，Young 和 Jones 所著的一个综合卷本(MYJ，或在微波行家中被简称为"黑皮书")中得到了很好的总结，对于那些对微波滤波器及与之紧密相关的阻抗匹配课题很感兴趣的人来说，这更是必读的。[25] 遗憾的是，这本大部头的书只出了一个版本，因此没有包括1960之后的一些进展。特别的是，与微带线实现有关的关键设计信息没有给出。

受图23.33中所示集总原型的启发，首先考虑电容性耦合谐振器，可以通过使用终端处的边缘电容来提供需要的耦合(或替代，或附加在分立的耦合电容上)。但是纯终端耦合传输线滤波器(参考图23.45)是很难搭建的，能够达到的耦合程度是受工艺上能够实现的最小间距限制的[26]。



图23.45　终端耦合的微带线带通滤波器

事实是，所需的空隙宽度不仅仅是不便实现的，它们通常小得不切实际。这个问题可以通过使用分立电容来补偿终端边缘得以很好的解决，同时减小了受尺寸的限制。然而，减少对分立元件的依赖通常是希望的，因而完全消除终端耦合的方法是值得考虑的。我们事实上已经在与耦合器设计的相关场合下连线中遇到过这个想法。于是，采用横向耦合作为替代就成了一种有力的解决方法[27]。

这个想法的一个简单的(而且是简化了的)但是实用的实现在图23.46给出。就如同终端耦合的滤波器，这些谐振器保持了标称为半波长的长度(这其中包括了对终端边缘效应的补偿)同时彼此之之间有四分之一波长的重叠[28]。具体的响应形状通过选择合适的线线耦合量以及这些传输线的特征阻抗来加以控制(为了简单起见，所有谐振器的线宽在图中都表示为相等)。

㉕ G. L. Matthaei, L. Young, and M. T. Jones, *Microwave Filters, Impedance-Matching Networks and Coupled Structures*(微波滤波器，阻抗匹配网络与耦合的结构)，McGraw-Hill, New York, 1964 (1980年由 Artech House 重新出版)。应该提一下这本书也涵盖了 Seymour Cohn 作出的重要贡献，他在用耦合传输线实现的滤波器上作出了广泛的、先驱性的工作，还包括一些其他的工作。Cohn 在斯坦福研究所(Stanford *R*esearch Institute, SRI)领导了一个团队，他们的工作形成了在 *MYJ* 这本书在给出的材料的大部分。

㉖ 然而，用分立电容器耦合的这些滤波器的版本很容易用类似 Puff 的工具来模拟，因此这些版本只有入门性的价值。

㉗ 再一次，参考 Seymou Cohn，但这一次是文献"Parallel-Coupled Transmission Line Resonator Filters"(平行耦合传输线谐振器的滤波器)，*IRE Trans. Microwave Theory and Tech.*, v. 6, April 1958, pp. 223-231。同时见 H. Ozaki and J. Ishii, "Synthesis of a class of stripline Filters"(带状线滤波器类的综合)，*IRE Trans. Circuit Theory*, v. 5, June 1958, pp. 104-109。

㉘ 这个交叠的值不是一个必须要求的，但因为这个值使得耦合最大化了，它使得相应的要求的间隔也最大化了，因而放宽了对尺度容差的要求。可以这么说，这是一个几乎普遍采用的选择。

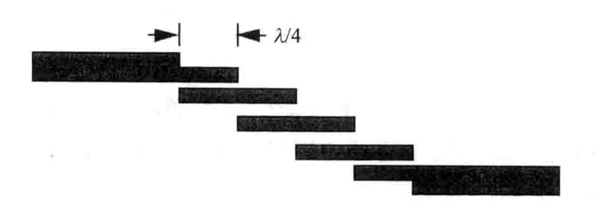


图 23.46 经典的边缘耦合微带线带通滤波器

为了设计一个显式的设计过程，我们直接利用这样一个认识，即耦合和阻纳逆变事实上是一件事，即便这个耦合并不是由于一个简单的机制形成的。因此让我们提议一对边缘耦合的传输线可以被建模为两个传输线由一个阻纳逆变器耦合在一起，如图 23.47 所示。

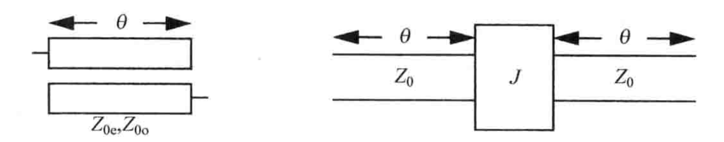


图 23.47 对称耦合线(左)以及建议的等效模型

如果两个网络是要等效的，一个要求是当它们的终端为相等的负载终断时，它们呈现出同样的输入阻抗。在研究传输线模型时(或连续或人为的)，我们已经看到一个无限长阶梯的输入阻抗计算是特别简单的。因此，尽管采用一个任意的负载阻抗在原理上是足够的，但如果计算一个无限级联的我们正在考虑的网络重复的输入阻抗，分析会大大地简化。我们通常将这个阻抗叫做网络的“特征”阻抗，但是为了避免与在系统中的各种传输线的特征阻抗相混淆，将无限重复网络的输入阻抗称呼为另外一个名字：镜像阻抗。

首先考虑图 23.48 中网络的镜像阻抗。这个镜像阻抗是频率的函数，但是让我们计算在中心频率的阻抗(其中 θ 为 $\pi/2$，对应于 $\lambda/4$ 线)。在中心频率的镜像阻抗(用下标 0 表示)因此容易由以下几个简单的步骤计算出：

$$Z_{I0} = \frac{Z_0^2}{\dfrac{1}{J^2(Z_0^2/Z_I)}} = J^2\frac{Z_0^4}{Z_I} \Rightarrow Z_{I0} = JZ_0^2 \tag{62}$$

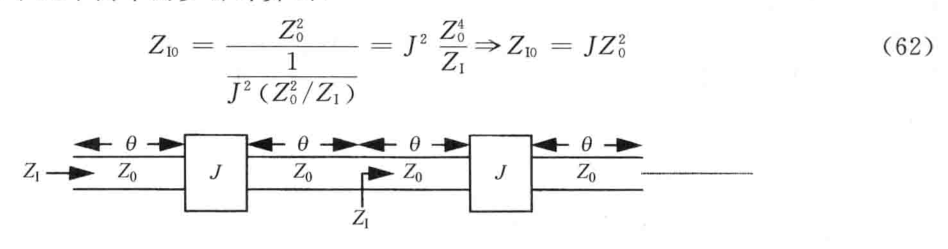


图 23.48 无限级联的传输线和逆变器

注意我们有意识地将第一个等式这样来表示以便于确定单个的阻抗逆变，而不是追求最简洁。

然而如果我们试图对级联耦合传输线的镜像阻抗直接进行计算，结果不会那么容易被推得。原因是这样的结构缺乏应用偶数模和奇数模分析所需的对称性(参考图 23.49)。

为了解决这个问题，我们间接地推导这个结构的镜像阻抗。首先用两端口参数对镜像阻抗推导一个通用的表达式，而后设法对这个耦合线系统找出这些参数的值。如果选择了正确的两端口表示，那么找到二端口网络参数的过程包括对称边界条件以能够进行偶数和

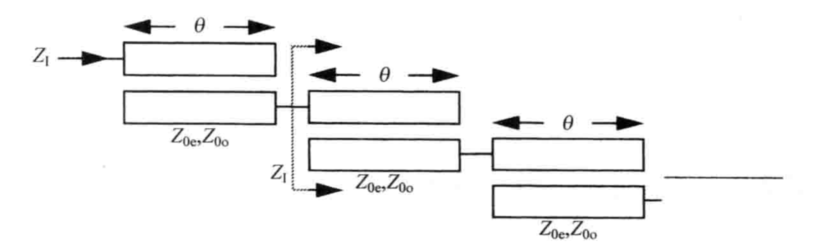


图 23.49　无限级联的耦合线

奇数模分解。具体地说，让我们考虑使用阻抗表示方式：

$$V_1 = I_1 Z_{11} + I_2 Z_{12} \tag{63}$$

$$V_2 = I_1 Z_{21} + I_2 Z_{22} \tag{64}$$

其中上面这些量如图 23.50 中所定义。

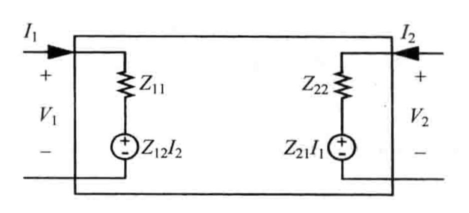


图 23.50　通用的阻抗两端口表示

现在考虑在输出端口两端连接一个其值为 Z_I 的阻抗，即镜像阻抗。对于这个对称网络，输入阻抗同样等于镜像阻抗。为了求解输入阻抗，可以（比如说）在输入端口加一个测试电压，计算流入电流的响应，而后求电压与电流的比。进行这些步骤就得到：

$$I_1 = \frac{V_{\text{test}} - Z_{12} I_2}{Z_{11}} \tag{65}$$

$$-I_2 = \frac{Z_{21} I_1}{Z_{22} + Z_I} \tag{66}$$

求解这些等式来得到 V_{test} 与 I_1 的比，有

$$\frac{V_{\text{test}}}{I_1} = Z_{11} - \frac{Z_{12} Z_{21}}{Z_{22} + Z_I} = Z_I \tag{67}$$

这给出了镜像阻抗的二次方程

$$Z_I^2 + Z_I(Z_{22} - Z_{11}) + Z_{12} Z_{21} - Z_{11} Z_{22} = 0 \tag{68}$$

我们初始的网络的对称性意味着其等效表示的端口参数也类似地有对称性，那么 $Z_{11}=Z_{22}$ 且 $Z_{12}=Z_{21}$。利用这些关系大大地简化了我们的任务，使得我们容易地发现：

$$Z_I^2 = Z_{11}^2 - Z_{21}^2 \tag{69}$$

让我们现在来解释这个结果。从对二端口模型的定义公式出发，可以看到 Z_{11} 为输出端口开路时测量到的输入阻抗。跨阻抗 Z_{21} 则是输出电压和输入电流的比，也是当输出端口开路条件下所测得的。幸运的是用于计算镜像阻抗的两个参数都是在开路条件下定义的，因此采用偶数和奇数模的分解就使得我们这个耦合传输线例子很快地得到参数，就如现在要演示的那样。

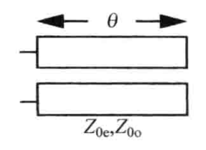


图 23.51　开路的耦合传输线

首先，在开路的条件下求解耦合传输线对的输入阻抗。如通常那样，将计算分解为偶数模和奇数模，分别进行计算，这是通过先用相同的电流，而后用等大但相反的电流来驱动两根线。参考图 23.51。

回顾一下，对于用开路终端的一根孤立的传输线，输入阻抗为

$$Z = -jZ_0 \cot\theta \tag{70}$$

那么类似地，对偶数模的激励有

$$Z_{11e}=-\mathrm{j}Z_{0e}\cot\theta \tag{71}$$

而对于奇数模激励，则有

$$Z_{11o}=-\mathrm{j}Z_{0o}\cot\theta \tag{72}$$

然后，总的开路输入阻抗就是这两个阻抗的平均值，

$$Z_{11}=-\mathrm{j}\left(\frac{Z_{0e}+Z_{0o}}{2}\right)\cot\theta \tag{73}$$

在中心频率上这个 Z_{11} 碰巧为零，因此我们现在不会用这个公式(但是之后将要用到)。

现在来考虑跨阻抗。同样地，从一根孤立的传输线开始。可以直截了当地证明一根终端开路的传输线的跨阻抗为

$$Z_{21}=V_2/I_1=-\mathrm{j}Z_0\csc\theta \tag{74}$$

注意，在中心频率下，跨阻抗的大小就是传输线的特征阻抗。

对于偶数模的激励，耦合传输线的跨阻抗为

$$Z_{21e}=-\mathrm{j}Z_{0e}\csc\theta \tag{75}$$

对于奇数模的激励，

$$Z_{21o}=-\mathrm{j}Z_{0o}\csc\theta \tag{76}$$

耦合传输线整体的开路跨阻抗因此为

$$Z_{21}=-\mathrm{j}\left(\frac{Z_{0e}-Z_{0o}}{2}\right)\csc\theta \tag{77}$$

在中心频率处($\theta=\pi/2$)求跨阻抗得到

$$Z_{210}=-\mathrm{j}\left(\frac{Z_{0e}-Z_{0o}}{2}\right) \tag{78}$$

因此，根据公式(69)，得到对于我们的耦合传输线频带中间的镜像阻抗为

$$Z_{I0}=\frac{1}{2}(Z_{0e}-Z_{0o}) \tag{79}$$

令这个镜像阻抗等于之前找到的逆变器模型的相应的频带中间值，

$$Z_{I0}=JZ_0^2 \tag{80}$$

给出我们一个方程。需要另外一个公式能够分别求解偶数和奇数模阻抗。有几种可能的选择，但是其中一个可以利用一些我们已经推得结果的选择是令两个网络的 Z_{11} 相等，这同样是在频带中间(或附近)来求值[29]。

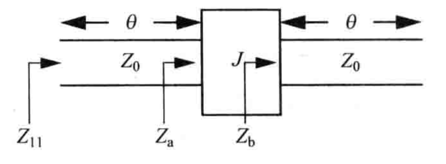


图 23.52 终端开路的等效逆变器模型

已经推导了耦合线的相关公式，所以只需推导图 23.52 逆变器模型中的 Z_{11}。从输出向输入倒推，有

$$Z_b=-\mathrm{j}Z_0\cot\theta \tag{81}$$

$$Z_a=\frac{1}{J^2Z_b}=\frac{\mathrm{j}\tan\theta}{J^2Z_0} \tag{82}$$

那么，

[29] 大多数处理方法是基于 Cohn 最早推出的结果，他将镜像阻抗与传输常数匹配起来。我们则选择将两个输入阻抗相匹配，因为这样做能简化推导过程。

$$Z_{11} = Z_0 \frac{\frac{Z_a}{Z_0} + \mathrm{j}\tan\theta}{1 + \mathrm{j}\frac{Z_a}{Z_0}\tan\theta} = Z_0 \frac{\frac{\mathrm{j}\tan\theta}{J^2 Z_0^2} + \mathrm{j}\tan\theta}{1 + \mathrm{j}\left(\frac{\mathrm{j}\tan\theta}{J^2 Z_0^2}\right)\tan\theta}$$

$$= \mathrm{j}Z_0\tan\theta \frac{\frac{1}{J^2 Z_0^2} + 1}{1 + \mathrm{j}\left(\frac{\mathrm{j}\tan\theta}{J^2 Z_0^2}\right)\tan\theta} \tag{83}$$

在中频带附近，分母中的正切值很大(在中心频率处则为无限大)，因此，可以将归一化输入阻抗近似为

$$Z_{11} \approx \mathrm{j}Z_0\tan\theta \frac{\frac{1}{J^2 Z_0^2} + 1}{\mathrm{j}\left(\frac{\mathrm{j}\tan\theta}{J^2 Z_0^2}\right)\tan\theta} = Z_0 \frac{\frac{1}{J^2 Z_0^2} + 1}{\frac{\mathrm{j}\tan\theta}{J^2 Z_0^2}} = Z_0 \frac{1 + J^2 Z_0^2}{\mathrm{j}\tan\theta} \tag{84}$$

将这个表达式与已推得的耦合传输线对应的公式比较：

$$Z_{11} = -\mathrm{j}\left(\frac{Z_{0e} + Z_{0o}}{2}\right)\cot\theta \tag{85}$$

这给出了所需的第二个公式：令它们相等，得到

$$\frac{1}{2}(Z_{0e} + Z_{0o}) \approx \frac{Z_0 \frac{1 + J^2 Z_0^2}{\mathrm{j}\tan\theta}}{-\mathrm{j}\cot\theta} = Z_0(1 + J^2 Z_0^2) \tag{86}$$

由两个中频带镜像阻抗相等，已经推断出：

$$\frac{1}{2}(Z_{0e} - Z_{0o}) = JZ_0^2 \tag{87}$$

加减公式(86)与(87)使我们最终能够求解每个模阻抗：

$$Z_{0e} \approx Z_0[1 + JZ_0 + (JZ_0)^2] \tag{88}$$

$$Z_{0o} \approx Z_0[1 - JZ_0 + (JZ_0)^2] \tag{89}$$

能够推出这些等式这个事实告诉我们，基于逆变器的模型的确是对我们最初的耦合传输线系统的一个近似表征。只需要记住在接近中频带处近似的程度较好，而距离中心频率越远，则近似的程度越差。因此，采用这些近似设计的滤波器，其通带性能与集总低通原型给出的预期相比很接近，但是阻带性能可能相差很大。

注意模频率的几何平均值并不等于 Z_0。然而，模阻抗的几何与算术平均值彼此在二阶精度内相等。

这里介绍如何应用我们推得的公式。可以计算针对集总原型带通滤波器的 J 逆变器的那些常数。这个知识，与谐振器的特征阻抗一起，使我们可以计算每对耦合传输的偶数和奇数模阻抗。最后，使用(比如说)Akhtarzad 公式来计算线宽和线间距，以产生给定的用微带线实现的偶数与奇数模阻抗。对于带状线，存在着对这些尺度的闭式解析表达式，可以用来简化设计。

边缘耦合带通滤波器设计过程总结

与逆变器耦合集总带通滤波器的推导相同，边缘耦合滤波器设计公式的推导也分散在几页上。为了使应用它们更方便，我们在此对设计过程做一个总结。幸运的是，使用公式比推导的过程要简单很多。

这些滤波器的设计再一次从集总原型的参数开始，而后的步骤与设计普通的集总带通滤波器相同。一个微小的差别是用作谐振器的传输线的特征阻抗与等效集总并联谐振回路有一些不同，就如我们已经在 $\lambda/4$ 谐振器看到的那样。对于一根末端开路的 $\lambda/2$ 传输线，使传输线与集总谐振器的导纳的导数相等导致了系数 $\pi/2$（参考 23.6 节）：

$$Z_{0\text{line}} = \frac{\pi}{2}\sqrt{L_\text{p}/C_\text{p}} = \frac{\pi}{2}Z_{0\text{res}} \tag{90}$$

为了顾及这个小的修正给出了接下来的一系列公式。其他部分的归一化逆变器参数与之前所给出的相同：

$$J_{j,j+1}Z_{0\text{res}} = \frac{(\text{BW})}{\sqrt{g_j g_{j+1}}} \Rightarrow J_{j,j+1}Z_{0\text{line}} = \frac{\pi}{2}\frac{(\text{BW})}{\sqrt{g_j g_{j+1}}} \tag{91}$$

单个谐振器的特征阻抗 $Z_{0\text{res}}$ 是一个自由变量，就如在集总谐振器情况下一样。例如，它可以被选择来最大化谐振器的 Q 值（来减小插入损耗以及与期望的响应形状的偏离）或最大化在谐振时的阻抗（这样可以使制造间隙时容易实现些）。这两个条件并不一定会同时得到满足，因此通常要做一些折中。损耗大一点的电介质如 FR4 倾向于使用比用高质量的材料制造的滤波器所用的线宽要窄些，后者中电导损耗占主导。尽管这么说，因为宽的（即低阻抗）传输线要求窄的线与线间距来达到给定的耦合，而窄的（即高阻抗）传输线会带来过高的电阻损耗，对任何一种工艺技术，谐振器的特征阻抗很少与 50～100Ω 的范围有所不同，因此上面所说的自由度只是在理论上存在。

就如对集总带通滤波器一样，可以在输入端和输出端使用逆变器将信号耦合进入及输出滤波器，同时又可提供任何必要的阻抗变换。相同的公式在此也适用（然而对谐振器与传输线阻抗的区别要做恰当的处理）。因此，假若继续将阻纳逆变器看作导纳，那么对于输入有

$$J_1 Z_0 = \sqrt{\frac{\pi}{2}\frac{(\text{BW})}{m g_1}} \tag{92}$$

其中 g_1 为原型低通滤波器中第一个元件的归一化值，而 m 为 $Z_{0\text{line}}/Z_0$。类似的，对于输出，

$$J_{n+1} Z_0 = \sqrt{\frac{\pi}{2}\frac{(\text{BW})}{m g_n g_{n+1}}} \tag{93}$$

其中 g_n 为最后一个原型元件的归一化值，而 g_{n+1} 为输出终端的归一化值。

然而，如果希望用耦合传输线来实现这些阻抗转化，一个复杂的情况会产生，因为我们仅对对称耦合传输线的情况推导了一个等效模型（见图 23.47）。在该模型中，阻纳逆变器不是一个孤立的元件，而是被两个特征阻抗相等的传输线段所包围。为了对最一般的情况（其中谐振器阻抗与系统阻抗不同）提供所需要的阻抗变换，逆变器周围的传输线要有不同的阻抗，一端是系统阻抗而另一端是谐振器阻抗。尽管出于这个目的可以使用非对称的耦合线，然而对它们的设计并没有简单的解析公式。因此，我们考虑用其他的替代选择。

首先注意到当 $m=1$ 时，一个对称传输线对就足够了，因此设计得以进行而不必考虑这个复杂性。这样，绕过这个非对称线问题的一个简单方法为选择所有线的阻抗都等于 Z_0。到目前为止，这个选择几乎是最常见的（常见到似乎不存在其他的选择）。

另外一个可能性是用一个任意的谐振器阻抗（比如说，同样选择能够最大化 Q 的阻抗）来进行设计，然后不用非对称的传输线对的方法来提供需要的匹配。例如，可以在每一端使

用 $\lambda/4$ 传输线来提供必要的匹配，或者是用恰当的集总阻纳逆变器。

比前面那些替代办法更简单的是对第一个和最后一个谐振器进行抽头降压来实现输入与输出端的耦合，因为阻抗从非常高的值（接近于 $\lambda/2$ 传输线的开路端）到几乎是零（接近于中心）连续地变化。这样的直接相连不再需要在输入输出端口设计和提供非对称的耦合器。同时，用抽头点的调整来产生最佳的匹配比起调整非对称耦合线的特征阻抗要容易得多。抽头降压主要的缺点是仅当谐振器阻抗超过 Z_0 的情况才有效。幸运的是，这个要求对几乎所有的实际例子都成立（而且可以人为使其成立）。

下一步用计算得到的逆变器的常数来推导所需要的耦合谐振单元的偶数和奇数模阻抗。从我们的推导中，结果可以被近似为

$$Z_{0e(j+1)} \approx Z_{0line}[1 + J_{j,j+1}Z_{0line} + (J_{j,j+1}Z_{0line})^2] \tag{94}$$

$$Z_{0o(j+1)} \approx Z_{0line}[1 - J_{j,j+1}Z_{0line} + (J_{j,j+1}Z_{0line})^2] \tag{95}$$

一旦求得每个谐振器的模阻抗，Akhatarzad 的方法（对微带线）可以用来确定谐振器的实际版图尺寸。输入与输出耦合结构的选择和实现最终完成了滤波器设计。在场求解器的指导下一些迭代可以给出必要的修正。

一个详细的设计例子

或许比本书中所讨论的其他滤波器更多的是，通过一步步对一个完整例子的设计来解释清楚设计过程。为了有利于与本章其他例子的比较，我们继续用有三个谐振器的 1dB Chebyshev 带通滤波器，其有波纹的通带从 950MHz 扩展到 1050MHz。使用几何平均值定义，我们的中心频率大约在 998.75MHz，以此为基础计算给出的归一化带宽（BW）大致为 0.1001。同样地，为了准确地说明过程，我们保留比实际所需的更多的有效数字。

首先找到对一个集总低通 Chebyshev 原型的归一化元件值为 $g_1=2.013$，$g_2=0.989$ 以及 $g_3=2.013$。在这个特别的例子中，使用软件 RFSim 99 中的快速滤波器设计工具来避免用前面一章中给出的公式计算元件参数的麻烦，或者在已发表的表中查得所需要的值。在很多情况下，许多滤波器设计表并没有对应于波纹如此大的 Chebyshev 滤波器的项。

下一步，选择将用作谐振器的传输线的特征阻抗。这里，假设 75Ω 能很好地平衡了低损耗和相对宽松的制造误差的要求。然后由以下公式可以计算在中带频率下的逆变器的导纳：

$$J_{j,j+1}Z_{0line} = \frac{\pi}{2}\frac{(BW)}{\sqrt{g_j g_{j+1}}} \tag{96}$$

对给定的数值，可以得到 $J_{12}=J_{23}=0.1114/Z_{0line}$，或大约 1.4858mS。因为滤波器的对称性，加上我们选择的较低的滤波器阶数，这些逆变器导纳是相等的。有相同终端负载的奇数阶，恒-k Butterworth 和 Chebyshev 滤波器是对称的。

偶数与奇数模阻抗可以容易地由下式计算得到：

$$Z_{0e(j+1)} \approx Z_{0line}[1 + J_{j,j+1}Z_{0line} + (J_{j,j+1}Z_{0line})^2] \tag{97}$$

$$Z_{0o(j+1)} \approx Z_{0line}[1 - J_{j,j+1}Z_{0line} + (J_{j,j+1}Z_{0line})^2] \tag{98}$$

因此，$Z_{0e2}=Z_{0e3}=84.3\Omega$ 并且 $Z_{0o2}=Z_{0o3}\doteq 67.6\Omega$。

下一步，考虑输入与输出匹配逆变器。源与负载阻抗都为 Z_0。因此，

$$J_1 Z_0 = J_4 Z_0 = \sqrt{\frac{\pi}{2}\frac{(BW)}{mg_1}} = \sqrt{\frac{\pi}{2}\frac{(BW)}{mg_N}} \tag{99}$$

对 $m=1.5$ 这里给出的值为 0.2282。这些阻抗匹配最好通过直接耦合来提供，这个直接耦合或者通过抽头，或者用 $\lambda/4$ 匹配单元来得到。我们后面会对这两个选择都加以考虑。

现在转向谐振器单元的实际版图尺寸计算。滤波器核心包括两个互相耦合的对，但是对称性允许将工作量减半。因此仅需要确定一个线宽，一个线间空隙，以及这些线的物理长度。

为了使读者避免将书翻来翻去的麻烦，这里再一次给出一小组在第 7 章已给出的 Akhatarzad 公式分。回想一下比例 W_e/H 与 W_o/H 是那些孤立的微带线传输线的值，其特征阻抗分别为 $Z_{0e}/2$ 与 $Z_{0o}/2$。在我们的例子中，$Z_{0e}/2$ 为 42.15Ω 而 $Z_{0o}/2$ 为 33.8Ω。使用众多将微带线尺寸与特征阻抗联系起来的公式中的一个，得到 W_e/H 大致为 2.38(对于体介质常数 4.6 的情况)而 W_o/H 大约为 3.31。用更细致的公式(例如在第 7 章中给出的参考文献)可以得到更好的精度，但是这里的目标是来说明整个设计过程，而不是使理论和实际的差别最小化。

在完成这个步骤之后，现在这步是多少更为困难些。我们迭代求解 Akhtarzad 公式来得到耦合线的实际宽度及间隙。这些公式如下：

$$\frac{W_e}{H}=\frac{2}{\pi}\operatorname{arcosh}\left(\frac{2d-g+1}{g+1}\right) \tag{100}$$

$$\frac{W_o}{H}=\frac{2}{\pi}\operatorname{arcosh}\left(\frac{2d-g-1}{g-1}\right)+\frac{4}{\pi(1+\varepsilon_r/2)}\operatorname{arcosh}\left(1+2\frac{W/H}{S/H}\right),\quad 如果\ \varepsilon_r<6 \tag{101}$$

其中，

$$g=\cosh\left(\frac{\pi S}{2H}\right) \tag{102}$$

$$d=\cosh\left[\pi\left(\frac{W}{H}+\frac{S}{2H}\right)\right] \tag{103}$$

来开始迭代的 S/H 的合理初值为

$$\frac{S}{H}\approx\frac{2}{\pi}\operatorname{arcosh}\left[\frac{\cosh\left(\frac{\pi}{2}\frac{W_o}{H}\right)+\cosh\left(\frac{\pi}{2}\frac{W_e}{H}\right)-2}{\cosh\left(\frac{\pi}{2}\frac{W_o}{H}\right)-\cosh\left(\frac{\pi}{2}\frac{W_e}{H}\right)}\right] \tag{104}$$

为了推断 W/H 的一个合理初值，注意到具有这些尺寸的传输线，其阻抗大致应该为具有尺寸 W_e/H 或 W_o/H 传输线的阻抗的两倍。考虑到阻抗与宽高比在零阶精度是成反比的，W/H 一个可行的初值为，比如说 $W_e/2H$。幸运的是，收敛性对初值的选择并不那么敏感，因此不必在具体使用那个初值上纠缠太多。

当对 Akhtarzad 方程进行迭代时，注意到这一点是有帮助的，即 S/H 主要控制 W_e/H 与 W_o/H 计算到的值之差，而 W/H 主要控制它们之和。这样利用该规律，可以大致正交地收敛于正确值：将 W_e/H 与 W_o/H 的正确值与给定 W/H 与 S/H 值条件下公式计算的结果比较，而后采取相应的对策。

遵循这些指导原则，在用数据表进行手工迭代一或两分钟之后，我们得到 $W/H=0.865$ 与 $S/H=1.066$ [30]。假设介质厚度为 1.6mm，线宽应当为 1.38mm 和线之间的间距

---

[30] 也可以用数据表的求解器的功能(这在一些数据表中是一个选项)来自动进行迭代。

为 1.71mm。同样地，注意我们提供的有效数字比实际有意义要更多(我们制造尺寸的重复性不大可能好到 10μm，而 Akhtarzad 方法并没有考虑到非零导电层的厚度)。

现在有了宽度和间距，需要来计算传输线的物理长度。首先由下式计算有效介质常数[31]

$$\varepsilon_{r,\text{eff}} \approx 1 + 0.63(\varepsilon_r - 1)\cdot(W/H)^{0.1255}, \quad W/H > 0.6 \tag{105}$$

继续假设一个体介质常数为 4.6，计算出有效介质常数为 3.227。在滤波器的中心频率 998.75MHz 下，半波长因此为 83.6mm。由于线两端处的边缘作用，版图的物理尺寸需要适当减小。通常的处理规则是在物理长度的每端添加 $H/2$ 得到估计电长度。然而，在这个例子中，终端边缘效应被临近线分走的通量有所减小。那么作为一个先验性的修正，对每个有邻近线的端做 $H/4$ 更正(即为传统修正的一半)。因此对中心的线选择 82.8mm 的物理长度，外侧两个线选择 82.4mm。由于我们把边缘效应的影响等效为长度的延长，版图也应该反映出相应的这种延长(参考图 23.53)。

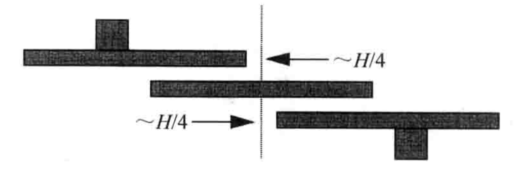


图 23.53 谐振器微带线的版图(画出了带有抽头的输入输出耦合)

为了完成第一轮的设计，需要设计出恰当的输入和输出匹配线段。如前面提到的，在输入和输出谐振器上直接在一个适当点处抽头有不少好处。在集总实现中，这个抽头可以在沿着电感的某一点实现或是在谐振器中使用抽头电容。阻抗在跨越并联谐振回路两端时为最大，其值为 $Q_L Z_{res0}$，其中 Q_L 为实际电路(即有负载)的 Q_L。而阻抗的最小值，即零，发生在接地端。因此可以有一个数值连续的阻抗。

一个不同阻抗的范围类似地可以沿着一个分布式振荡器得到。阻抗在两个终端开路的端处最大(同样地其值为 $Q_L Z_{res0}$)，而在 $\lambda/2$ 谐振线段的中心处为最小(~0)。如果假设电流和电压沿谐振微带线近似地以正(余)弦形式变化，那么阻抗应该以一种与类似于贴片天线的嵌入馈电点的阻抗方式变化。形式上，但是近似地[32]，

$$(\sin\theta)^2 \approx \frac{\pi}{2}\,\frac{Z_0}{Q_L Z_{res0}} \tag{106}$$

其中 Q_L 是终端谐振器的(双倍)加了负载的 Q，

$$Q_L = \frac{g_1}{2(\text{BW})} \tag{107}$$

我们最终要求解从谐振器的中心到抽头点的距离 θ(以相角的形式表示)，在抽头点滤波器对外电路表现出 Z_0 的阻抗。对我们所用的数值，Q_L 大致为 20.1，谐振器阻抗为 75Ω，而我们要匹配的系统为 50Ω。因此 θ 应当为 0.33rad，对应于计算得到的抽头位置大约为 8.5mm(这是相对于谐振器中心位置而言)。这个位置可以用作求解抽头最佳点的迭代初值。

在我们的情形下，应用场求解器的试探揭示了抽头点的最佳位置实际上是距离谐振器中心 11mm 处，参见图 23.54。用 Sonnet Lite 9.51 对设计进行仿真表明存在着一定的插入

---

[31] 就如与微带线特征阻抗的计算，存在着大量的计算有效电介质常数的公式。

[32] 见 *Reference Data for Radio Engineers*(无线电工程师参考数据手册)，5th ed.，International Telephone and Telegraph Corp.，1969。也可参考 Joseph S. Wong，"Microstrip Tapped-Line Filter Design"(带抽头的微带线滤波器设计)，IEEE Trans. *Microwave Theory and Tech.*，*January* 1979，pp. 44-50.

损耗(1.4dB 的量级——这可能是由于辐射的发生，由于滤波器是被由无损导体所组成的墙环绕的，在仿真中只有顶部开放到自由空间)。重新归一化到这个损耗，波纹的峰峰值大约为 1.2dB，而有波纹的通带从 940MHz～1.04GHz。因此，波纹只超过设定的目标一点点，中心频率也类似地小一点(大约 1%)，而 100MHz 的波纹带宽是与期望相同的。如意料之中，通带比过渡带和阻带更接近于期待值。尽管有这些偏差，整体的性能对于第一轮设计来说是令人满意的，特别是鉴于到达这一地步我们已经做了很多近似。

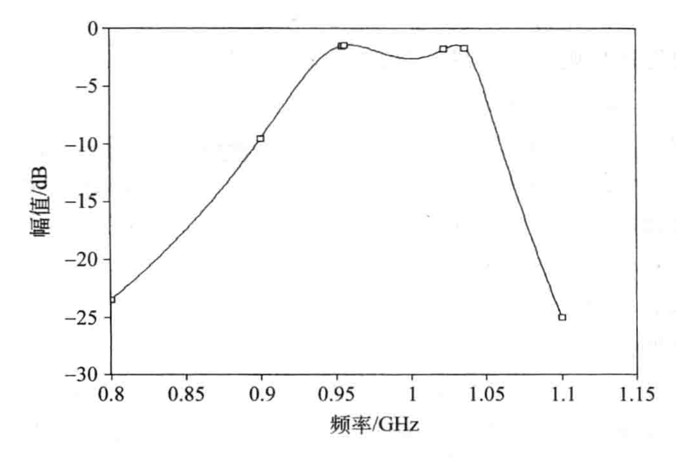


图 23.54　仿真得到的滤波器响应(采用软件 Sonnet Lite 9.51)，上边界条件是自由空间，在 11mm 处抽头

作为对设计步骤的最后一个评估，我们对采用略为不同的边界条件来重新进行仿真。现在为包住滤波器的盒子加一个用理想导电的导体做成的顶盖，而在滤波器导体上方有 100mm 空气作为电介质。在这种情形下，插入损耗的峰值的确下降到接近为零，这是因为辐射是预先排除掉了。然而，波纹增加到 2.5dB。这个 2.5dB 波纹通带从 950MHz 扩展到大约 1054MHz，而 1dB 的边缘发生在 954MHz 与 1051MHz。因此，取决于我们如何来解释这些数值，这个带宽或者是稍高一点，或者是略低一些。无论何种情况，中心频率看上去是略为高了些。

除了那些改变外，当抽头位置向中心更靠近些会实现最好的输入输出匹配。对这个特别的例子中，最佳的抽头点大约距离中心 9mm，如图 23.55。

通过改变谐振器之间的间距可以减小波纹。这个大的波纹值意味着邻近的谐振器间的耦合是过于紧密了，以至产生过大的模分裂。因此，进一步迭代可以尝试将谐振器的间距再略微增大些，然后根据需要调整抽头的位置，从而在整个过程中保持良好的匹配。

主题的变化

边缘耦合带通滤波器应用十分广泛，这是因为它们设计起来相对比较直截了当且性能也很好。但是一种批评认为这种滤波器占据了相对大了点的面积(或者其形状不规则)，特别是当滤波器需要使用很多段时。为了更佳地填充一个矩形空间，通常可以将谐振器阵列的版图加以旋转，如图 23.56 所示。输入输出耦合传输线通常是被弯曲了的(如图 23.56 所示)而端口被保持在水平方向上放置。这类滤波器的设计过程与传统的边缘耦合滤波器完全相同。旋转仅仅以二阶影响性能，因而所需的修正也十分小。

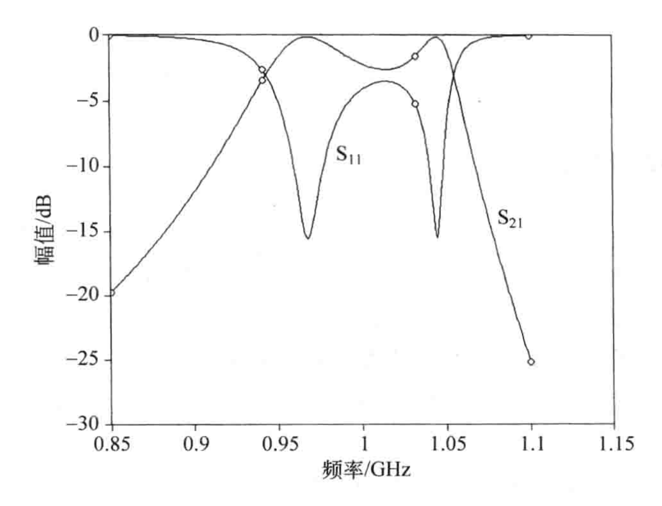


图 23.55　带有理想导电顶盖结构的仿真响应；抽头在 9mm 处



图 23.56　更为紧凑的边缘耦合带通滤波器版图

为了进一步缩小滤波器，可以将每个单独的谐振器折叠成发夹的形状[33]，如图 23.57 所示。注意如其他弯折传输线一样，谐振器段使用斜弯来使不连续性最小化，因为不连续性是会改变频响的。



图 23.57　发夹带通滤波器(输入-输出耦合没有给出)

在一个给定的发夹中的平行臂之间的耦合在一些情况下被证明会给这一类滤波器带来麻烦，特别是当倾斜角(或称倾斜系数)α 比较小时。在发夹臂两个相反端上的电压是反相的，因此某种类似于经典密勒效应的作用使得臂与臂之间的电容出现不成比例的效应。这个增加了的电容降低了谐振频率，需要发夹长度的缩短来进行补偿。为了最小化这个自耦合的问题，应当通过增加倾斜系数尽可能地使臂与臂的间距加大。然而，因为需要减小谐振器线的线间距来补偿变小了的横向的交叠，在可实现的臂间间距有实际的限制。一般说来，五倍于电介质厚度的臂间间隔是两个相悖的要求之间的良好折中。

对发夹形滤波器的设计并没有简单的解析公式存在，因此大多数工程师用迭代的方法进行设计，同时由场求解器或用能够将过程自动化的 CAD 工具将作为指导。然而，可能设计一个近似的方法为这种迭代提供一个合理的初值。为了这样做，首先重复两根线间耦合的一些公式：

---

[33] E. G. Cristal and S. Frankel, "Hairpin Line/Half-Wave Parallel-Coupled-Line Filters"(发夹形传输线-半波平行耦合线滤波器)，*IEEE Trans. Microwave Theory and Tech.*, v. 20, November 1972, pp. 719-728。对一个改进了设计步骤，见 U. H. Geysel, "New Theory and Design for Hairpin-Line Filters"(发夹形传输线滤波器的新理论与设计)，IEEE Trans. Microwave Theory and Tech., v. 22, May 1974, pp. 523-531。

$$\frac{V_{\text{cpld}}}{V_{\text{in}}}=\frac{\mathrm{j}[\tan\theta]\dfrac{Z_{0\mathrm{e}}-Z_{0\mathrm{o}}}{Z_{0\mathrm{e}}+Z_{0\mathrm{o}}}}{\dfrac{2Z_0}{Z_{0\mathrm{e}}+Z_{0\mathrm{o}}}+\mathrm{j}\tan\theta}=\frac{[\mathrm{j}\tan\theta]C_{\mathrm{F0}}}{\sqrt{1-C_{\mathrm{F0}}^2}+\mathrm{j}\tan\theta} \tag{108}$$

在中频带，这个比例可简化为

$$\frac{V_{\text{cpld}}}{V_{\text{in}}}=C_{\mathrm{F0}}=\frac{Z_{0\mathrm{e}}-Z_{0\mathrm{o}}}{Z_{0\mathrm{e}}+Z_{0\mathrm{o}}} \tag{109}$$

尽管这些公式严格说来仅对有端接的耦合线成立——而不是那些在我们的滤波器中有端接的传输线——我们将以一种使其最后结果对这种区别不敏感的方式来使用这些公式。在对一个普通的边缘耦合滤波器求得偶数和奇数模阻抗的组合之后，对于每一对耦合传输线，用公式(109)来计算比例 $V_{\text{cpld}}/V_{\text{in}}$。注意我们已经用下标 0 来表示适用于原型滤波器的耦合系数。

接下来，计算包含倾斜系数 α 时，产生相同 $V_{\text{cpld}}/V_{\text{in}}$ 大小的 C_{F} 值：

$$\left|\frac{V_{\text{cpld}}}{V_{\text{in}}}\right|=\left|\frac{[\mathrm{j}\tan(\theta-\alpha)]C_{\mathrm{F}}}{\sqrt{1-C_{\mathrm{F}}^2}+\mathrm{j}\tan(\theta-\alpha)}\right| \tag{110}$$

令在中带频率下对两个情况的比例 $V_{\text{cpld}}/V_{\text{in}}$ 大小值相等，得到：

$$C_{\mathrm{F0}}=\left|\frac{\left[\mathrm{j}\tan\left(\frac{\pi}{2}-\alpha\right)\right]C_{\mathrm{F}}}{\sqrt{1-C_{\mathrm{F}}^2}+\mathrm{j}\tan\left(\frac{\pi}{2}-\alpha\right)}\right| \tag{111}$$

因此，首先如同设计一个普通的倾斜系数为零的边缘耦合滤波器那样来计算 C_{F0}。然后解方程(111)(比如，采用迭代的方法)来得到对给定的非零倾斜角所需要的 C_{F} 值。那个计算得到的(更高的)耦合系数被用来从下式推导新的偶数和奇数模传输线阻抗：

$$\frac{Z_{0\mathrm{en}}-Z_{0\mathrm{on}}}{Z_{0\mathrm{en}}+Z_{0\mathrm{on}}}=C_{\mathrm{F}} \tag{112}$$

其中下标 n 表示“新”的意思。

还需要另外一个方程来计算单个模阻抗。作为第一个近似，可以假设这两个模阻抗的乘积与倾斜系数无关。这个额外的关系就足够用来计算新的模阻抗，在这之后可以确定为发夹振荡器系统需要的需要的传输线及其间隙的尺寸(例如，同样是使用 Akhtarzad 方法)。

这样计算得到的尺寸甚至比不采用这个方法得到的更为近似，这是因为我们已经描述的过程并没有考虑到每个发夹的臂之间的耦合。尽管这么说，实际经验表明采用这种方法通常能够得到令人满意的或是经过一些小修改就可令人满意的第一轮设计。

就如与其他很多现象一样，在一个场合下是造成麻烦的事，换一个场合却会变成可以加以利用的。在这里，臂之间的耦合可以认为是有利条件，因为它提供了进一步减小尺寸的机会。一旦我们的思路趋向于这个方向，我们的目标就转为考虑如何去增强这个效应，而不是减弱。这种思考方式的一个结果就是折叠的发夹滤波器，如图 23.58 所示(输入-输出耦合没有显示出来)[34]。

---

[34] M. Sagawa, K. Takahashi, and M. Makimoto, “Miniaturized Hairpin Resonator Filters and Their Applications to Received Front-end MIC’s”(微型发夹谐振器滤波器及它们在接收器前端微波集成电路上的应用), *IEEE Trans. Microwave Theory and Tech.*, v. 37, no. 12, December 1989, pp. 667-670。

图 23.58　折叠发夹带通滤波器(尺寸没有严格按照比例)

那些在每个谐振器内部的折叠单元可以当作是形成了横向电容。它们在臂之间产生了一个增强了的电容耦合，从而降低了谐振频率，这样就允许(事实上，是要求)滤波器尺寸的减小。如可以预计到的那样，折中是不存在简单的解析公式可以用来指导它们的设计。通常我们从发夹原型开始，而后用场求解器来计算折叠对共振频率的影响。然后，谐振器之间的间隙连续地被减小来保持所希望的耦合。继续这个过程直到得到一个满意的设计，或者是你对再尝试感到厌倦了。注意也可以使用跨接臂末端间的集总电容，这又是在同一主题上做出的另一种可能的变化。

这个折叠发夹滤波器仅仅是过去几十年间发展出来的许多变种之一。为了强调创意从未被穷尽，再给出两个结构，但不做分析。

其中之一为叉指式滤波器(要比折叠发夹设计早几十年)。可以将这个滤波器想象为将一个传统边缘耦合设计中每个 $\lambda/2$ 谐振器自身折叠的结果。这种折叠产生了一系列互相耦合的 $\lambda/4$ 谐振器滤波器，如图 23.59 所示。那些设计方程与 $\lambda/2$ 边缘耦合滤波器的设计方程本质上是相同的。因为在电抗斜率上有 2 倍的因子的差别，需要作些小的修正，但是除了这个小细节以外，设计的过程与之前一样。

叉指式滤波器要求有接地的连接，并且是在谐振器的交替端实现的，这非常不方便，因此这种滤波器并不以微带线形式广泛使用。尽管如此，一旦拓宽思路来考虑接地的连接，就值得来考虑谐振器的另一种构造方法，即所有的接地连接都在一端，如图 23.60。

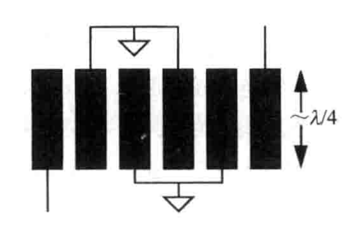


图 23.59　叉指型滤波器

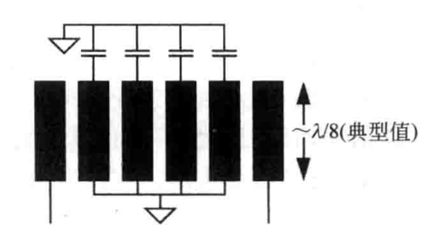


图 23.60　梳状线滤波器

梳状线滤波器的谐振器通常如图中所示的来实现，其中线段的长度短于 $\lambda/4$(事实上，典型值为 $\lambda/8$)，而且与分立电容一起使用。使用电长度如此小的传输线使得寄生通带不容易出现，同时也缩小了滤波器整体大小。另外，使用分立电容有利于调谐。最后是，对带损耗的介质(例如 FR4)，总的并联谐振回路 Q 值因为使用质量好的分立电容器作为谐振器的一部分而得以提升，因此就降低了插入损耗(而且同时减小了理论期望和实际性能之间的差别)。采用这个结构可以达到通常只有非常新颖(因此昂贵)的介质板材才能实现的高性能。这些优势有时足以使梳状线滤波器十分有吸引力，尽管在接地要求上有不便之处。

23.4 实际的考虑

当试图去构建一个实际滤波器时，每个滤波器设计者都会遇到一些严重的实际困难。其中之一是适应制造公差。滤波器性能特征越重要，在设计、构建和调谐电路时就越需要仔细——这一点通过仔细分析在前面那一章讨论过的许多 Monte Carlo 仿真就可以看出。对于要求高的应用，我们已经给出的相对简单的步骤通常不能给出满足所有指标的设计。对这些设计的进一步改进最好有直觉指引，然后用严格的工具，例如场求解器，来加以验证[35]。

另外一个挑战是开发一个行得通的步骤，以便在滤波器造好之后就能对其进行调正。对低阶滤波器，为设计添加足够的余量来容纳制造的公差并不需付出太多的代价，因此不需要再进行调正。然而对于要求高的应用，由于各种原因，上述选择可能无法接受，我们不得不来设计并实现一个调正过程。

Dishal 在 1949 年所描述的一个聪明的对准方法的修改版本对任何耦合谐振器滤波器（从理论上）都很有效，不论这个滤波器是用集总或分布式形式实现的。这个方法同时也利用了一些我们之前作出的一些观察[36]。其中一个观察是孤立回路的谐振频率都是相等的。另一个则是，耦合网络（J 和 K 阻纳逆变器）近似为 $\lambda/4$ 线。为了看看只用这两个结论我们能够走多远，考虑在对并联谐振回路调频时输入阻抗是如何变化的。一个代价较低的方式是将信号产生器的频率设到所要求的中心频率，而后在它与滤波器输入之间放置一根开槽测试线。这个开槽线不需要被校准，所以如第 8 章中描述的手工做的就适合了。然后，将所有滤波器并联谐振回路都完全调偏谐振频率（例如，使谐振并联谐振回路两端尽可能好地进行短路）。移动开槽线的探针直到找到最小值点并在后面的调正过程中将其锁定在这个位置。调整第一个谐振器（即距离输入端最近的那一个）以给出探测读数的峰值（这是由对应的阻抗峰值造成的）。然后调整下一个谐振器来给出最小的响应。这个最小值之所以产生是因为当第二个谐振器被正确地调整后，它的阻抗是一个最大值。当通过阻纳逆变器的类似于 $\lambda/4$ 线的变换，输入阻抗就变成最小。稍加思考就可以得到这样的结论：当调整奇数编号的并联谐振回路时我们寻求得到最大的探针读数，而当调整偶数编号的并联谐振回路时寻求最小值的读数，就可以达到调谐目标。这个方法之所以有吸引力，是因为不需要对滤波器进行二端口表征，且只需要对 n 个谐振器的滤波器进行 n 次调整。唯一需要校准的仪器是一个（固定频率的）信号发生器。

前面的过程保证了谐振器得到恰当的调整，但是我们还没有完全完成滤波器的对准，因为依然必须调整谐振器之间的耦合系数。在前面关于模分裂的讨论中提出的步骤可以用作

---

[35] 这些工具以往的价格远非是学生和业余爱好者所能承受的。幸好，近年来这种情况大为改观，因为有一些软件厂商提供了他们的场求解器的免费或便宜的学生与演示版本。有代表性的是 Sonnet Lite（包括在与本书一起提供的 CD-ROM 集锦），它是一个免费的可以进行相当多的先进分析的电磁场求解器。用这个工具，可以直截了当地以相当高的精度来包括边缘效应、耦合与其他效应，这样就使得微带线滤波器的设计要比先前的状况容易得多。这些程序的广泛使用正在引起学生与业余爱好者这一类人的工程实践的革命性变化。甚至是周末的动手的爱好者现在都能很快地完成相当复杂的设计，这在仅仅是几年前都是不可想象的。

[36] Milton Dishal, "Alignment and Adjustment of Synchronously Tuned Multiple-Resonant-Circuit Filters"（同步调谐的多个谐振电路滤波器的对准与调整），*Proc. IRE*, November 1951, pp. 1448-1455，及"Alignment and Adjustment of Synchronous Tuned Multiple-Resonant-Circuit Filters", *Electrical Commun.*, June 1952, pp. 154-164。

这种调整的基础。在耦合被适当确定之后，来重新检查谐振器的调整是个好主意。由于谐振器之间的耦合通常是较小的，对它们的调整而引起的对谐振回路调谐的改变通常是可以忽略的。然而，最好总是对这个预期进行验证，特别是当在前面的步骤中对耦合需要进行大的调整或滤波器的指标要求太高。即使需要对一些谐振器重新调整，耦合系数很少需要第二次调整迭代。无论在哪种情况下，用这种技术得到合适的对准的收敛过程是非常快的，因此在实践中只需很少数几次迭代就可以完成这个过程。

另一个重要的观察是所有我们的设计公式和过程对所有的元件都假设了无穷大的空载 Q。对其极点的 Q 值远小于元件空载 Q 的滤波器，偏差不是那么严重。然而随着滤波器阶数和复杂度的增加，对高 Q 极点（和零点）的需要也增加了，使得寄生造成的损耗的问题越来越变得严重。插入损耗将增加，通带精度会下降，而且应该是传输零点的频率处也不再有无限大的衰减。采用更现代化的综合方法，在设计过程中可以兼顾某些损耗（例如，通过改变 L/C 的比例以使得电路和自身负载的组合最终产生一个与对无损元件设计得到的相同的带有负载的 Q），但是这一切是以增加插入损耗为代价的[37]。很遗憾，对传输零点我们就没有办法了。在所有情况下，对任一个滤波器都存在某个临界的 Q 值，低于这个值不可能得到可接受的滤波器特性。对这些情况下，唯一的那些备选方法是用更好的工艺技术来实现滤波器。在这个意义上，不应忽略使用集总和分布式元件组合（例如梳状线架构）的可能性。

23.5　总结

我们已经看到，在集总电路范畴里发展起来的一些内在见解可以挪用到分布式范畴，使得我们能够理解很多重要微带线滤波器的工作原理。通常，甚至可以从集总原型的公式中推导出解析设计公式来。

我们也看到在许多耦合线滤波器的谐振单元中产生的相对高的电场，意味着对 FR4 不算太高的介质损耗的特别的敏感性。将一些带损耗的线用低损耗分立电容来取代可以提升性能，正像在电容调谐的梳状线滤波器中那样。出于这个原因，如果要使用相对高损耗的材料如 FR4，关键的滤波器（和谐振器）最好通过这类混合微带线——分立元件的组合来实现。

23.6　附录　分布式谐振器的集总等效形式

在很多设计方法中，我们将一根谐振传输线的特性与一个等效集总谐振器的那些特性相匹配。为了这样做，在接近谐振处将这两个网络的行为进行匹配。具体地说，在共振频率处匹配电抗（或导纳）的斜率。照例，从一根带负载的传输线的输入阻抗表达式开始：

---

[37] 这个方法在文献中被称为预失真。在 Zverev 的书（*Handbook of Filter Synthesis*，滤波器综合手册，Wiley，New York，1967）中的丰富的以表格形式出现的设计包括了通过预畸变来兼顾有限 Q 值元件的滤波器。再一次，我们要感谢 Dishal，是他首次给出了对预畸变的显式公式，这是在“Design of Dissipative Band-pass Filters Producing Desired Exact Amplitude-Frequency Characteristics”（能产生所要求的准确幅度-频率特性的带有损耗的带通滤波器设计），*Proc. IRE*，September 1949，pp. 1050-1069。

$$\frac{Z(z)}{Z_0} = \frac{\frac{Z_L}{Z_0} + \mathrm{j}\tan\beta z}{1 + \mathrm{j}\frac{Z_L}{Z_0}\tan\beta z} \tag{113}$$

对于一根开路线，有

$$\frac{Z(z)}{Z_0} = \frac{1}{\mathrm{j}\tan\beta z} \tag{114}$$

为了便于与集总并联谐振回路的比较，实际上将考虑导纳，

$$\frac{Y(z)}{Y_0} = \mathrm{j}\tan\beta z \tag{114}$$

现在，

$$\beta = \omega/v \tag{116}$$

并且，

$$\lambda = 2\pi v/\omega \tag{117}$$

因此，一根在中心频率处为 $\lambda/2$ 长的线，其长度可以表示为

$$\lambda_0/2 = l = \pi v/\omega_0 \tag{118}$$

这样，

$$Y(l) = \mathrm{j}Y_0 \tan\frac{\omega\pi}{\omega_0} \tag{119}$$

现在让我们计算在中心频率处导纳的导数：

$$\left.\frac{\mathrm{d}Y}{\mathrm{d}\omega}\right|_{\omega=\omega_0} = \mathrm{j}Y_0\frac{\pi}{\omega_0} \tag{120}$$

对一个并联 LC 谐振回路进行同样的计算，得到：

$$\left.\frac{\mathrm{d}Y}{\mathrm{d}\omega}\right|_{\omega=\omega_0} = \mathrm{j}2C \tag{121}$$

令两个导纳的斜率相等就得到了已经反复使用的关系：

$$Y_0(\pi/\omega_0) = 2C \Rightarrow Z_0 = \frac{\pi}{2}\sqrt{L/C} = \frac{\pi}{2}Z_{0\mathrm{res}} \tag{122}$$

就是说，一个特征阻抗为 Z_0 的 $\lambda/2$ 传输线表现为一个阻抗是 $(2/\pi)Z_0$ 的并联 LC 谐振回路。